بسم الله الرحمن الرحیم

آمار و مدل‌سازی

سال دوم آموزش متوسّطه

رشته‌های ریاضی و فیزیک ـ ادبیّات و علوم انسانی ـ
علوم و معارف اسلامی

و

سال سوم آموزش متوسّطه

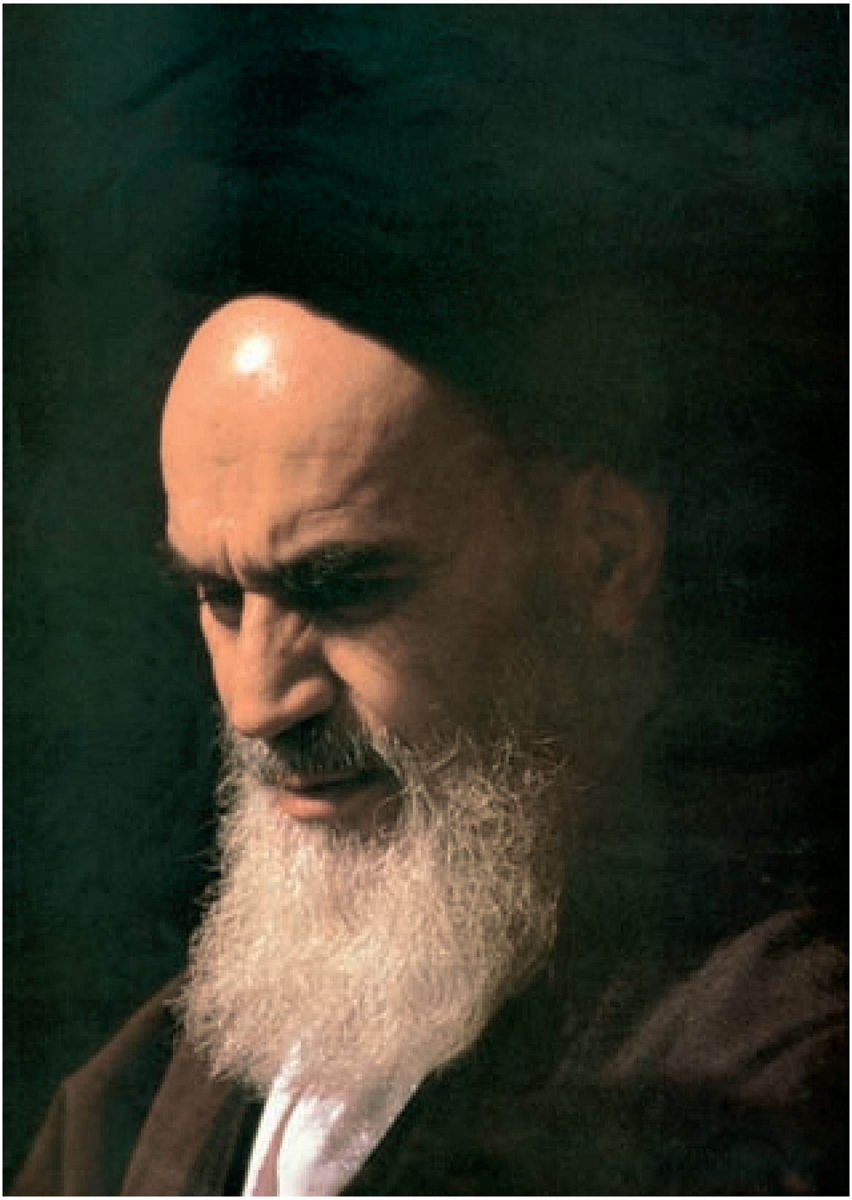
رشتهٔ علوم تجربی

۱۳۹۱

وزارت آموزش و پرورش
سازمان پژوهش و برنامه‌ریزی آموزشی

برنامه‌ریزی محتوا و نظارت بر تألیف : **دفتر برنامه‌ریزی و تألیف کتاب‌های درسی**
نام کتاب : **آمار و مدل‌سازی ـ ۲۳۴/۳ و ۲۵۸/۵**
مؤلفان : **شهرناز بخشعلی زاده، عین‌الله پاشا و آرش رستگار**
آماده‌سازی و نظارت بر چاپ و توزیع : **ادارهٔ کل چاپ و توزیع کتاب‌های درسی**
تهران: خیابان ایرانشهر شمالی ـ ساختمان شمارهٔ ۴ آموزش و پرورش (شهید موسوی)
تلفن: ۹ـ۸۸۸۳۱۱۶۱، دورنگار: ۸۸۳۰۹۲۶۶، کدپستی: ۱۵۸۴۷۴۷۳۵۹،
وب‌سایت: www.chap.sch.ir
رسام : **هدیه بندار**
صفحه‌آرا : **خالد قهرمانی‌دهبکری**
طراح جلد : **محمّدحسن معماری**
ناشر : **شرکت چاپ و نشر کتاب‌های درسی ایران: تهران ـ کیلومتر ۱۷ جادّهٔ مخصوص کرج ـ خیابان ۶۱ (داروپخش)**
تلفن: ۵ ـ ۴۴۹۸۵۱۶۱، دورنگار: ۴۴۹۸۵۱۶۰، صندوق پستی: ۱۳۹ـ ۳۷۵۱۵
چاپخانه : **شرکت چاپ و نشر کتاب‌های درسی ایران «سهامی خاص»**
سال انتشار و نوبت چاپ : **چاپ سیزدهم ۱۳۹۱**

شابک ۵-۰۷۳۸-۰۵-۹۶۴ ISBN 964-05-0738-5



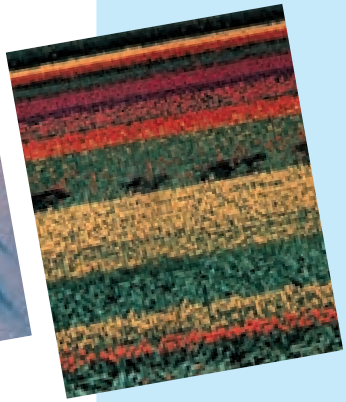
امام به بحث و مذاکره در مسایل علمی و بحث و مذاکره در مجلس درس علاقهٔ فراوانی داشتند. اگر برجلسهٔ درس سکوت حکمفرما می‌شد، می‌فرمودند : مجلس درس است، نه مجلس روضه. سخن بگویید و اعتراض کنید.

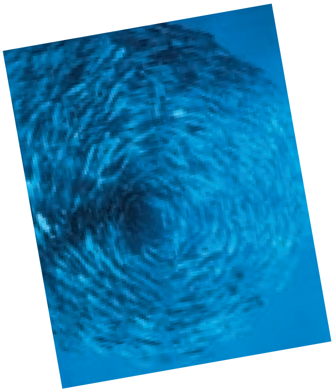
فهرست













سخنی با خواننده

می‌گویند هستی موضوع هیچ علمی به‌تنهایی نیست. این گفته بدین معنی است که پرداختن به زمینه‌های یک علم، فقط گوشه‌های بسیار کوچکی از هستی را برای ما روشن می‌کند. برای آن‌که دامنهٔ شناخت وسعت یابد، لازم است چند شاخهٔ علمی به هم گره بخورد و از روش‌های یکی، در دیگری استفاده شود. ریاضیات به‌عنوان یک علمِ پایه‌ای و زیربنایی حضور خود را در بسیاری از علوم به اثبات رسانده است. آمار نیز از یک دیدگاه به‌عنوان علمی مستقل و از دیدگاهی دیگر به‌عنوان شاخه‌ای از ریاضیات، در علوم کاربردی جایگاه ویژهٔ خود را دارد.

به‌علاوه در دنیای کنونی، سرعت دگرگونی به‌حدی است که **عمر مفید اطلاعات به یک دهه هم** نمی‌رسد. بنابراین انباشتن ذهن با اطلاعاتی که به‌سرعت رو به زوال است جز اتلاف وقت و انرژی حاصلی نخواهد داشت. به همین علت است که در برنامه‌های اخیر دولت، به جای **انتقال دانش، بر تولید دانش** تکیه شده است. برای آن‌که بتوانیم به این مقصود برسیم، لازم است روش‌ها و عادت‌های آموزشی خود را **تغییر** دهیم و به روش‌هایی توسل جوییم که ما را در رسیدن به تولید دانش و شناختِ بهتر از محیط اطراف، کمک کند. یکی از اساسی‌ترینِ این روش‌ها آن است که سطح توانایی خود را در **مدل‌سازی** پدیده‌های مورد مطالعه افزایش دهیم. مدل‌سازی به بیان ساده یعنی بیان مسئله و پدیدهٔ مورد مطالعه به زبان یک علم.

هدف عمومی این درس ایجاد زمینه‌هایی در دانش‌آموزان است که آن‌ها را قادر سازد علاوه بر انجام اعمال ریاضی کاربرد ریاضیات را در زندگی روزمره مشاهده نموده و به بررسی مسائل محیط اطراف بپردازند. معمولاً مسائلی که در آمار و مدل‌سازی مطرح هستند به گونه‌ای هستند که افق‌های تازه‌ای را در برابر دانش‌آموزان باز می‌نمایند و در نتیجه دانش‌آموزان از وسعت دید بیش‌تری برخوردار می‌شوند. به‌طورکلی این درس توانایی‌های زیر را در دانش‌آموزان تقویت می‌نماید :

۱ــ پرورش قدرت مشاهده

۲ــ افزایش دقت دانش‌آموزان و عادت به یک نظم فکری

۳ــ پرورش قدرت استدلال و استفاده از تکنیک‌های آماری به‌عنوان روش‌های استدلال و تفسیر

۴ــ پرورش قدرت تجزیه و تحلیل داده‌ها و فرضیه‌ها و تشخیص مناسب بودن آن‌ها

۵ــ پرورش تفکر نقاد و قدرت بیان و ارائه نتایج حاصل از استدلال و تفسیر

۶ــ پرورش روحیه‌ی همکاری و توانایی‌ها و قابلیت‌های زبانی و مهارت‌های برقراری ارتباط در زندگی اجتماعی و یادگیری‌های بعدی

۷ــ پرورش قدرت حل مسائل در ارتباط با محیط اطراف

۸ــ آشنایی با ابزار تکنولوژی و استفاده از آن‌ها در آموزش

قسمت‌هایی از کتاب که با ترام صورتی مشخص شده برای مطالعه در نظر گرفته شده است و جزو برنامه اجباری نمی‌باشد. همچنین فصل هشتم به صورت آزمایشی به کتاب افزوده شده است.

معلمین با در نظر گرفتن توانایی دانش‌آموزان و در دست داشتن وقت کافی می‌توانند این فصل را در برنامه تدریس خود بگنجانند.

من پرچم هدایتم و پناهگاه پرهیزگاری و جایگاه سخاوت و دریای جود و جوانمردی کوه خرد. من صدّیق اکبرم و علم خدا هستم و قلب آگاه خدایم و زبان گویای خدا و چشم خدا و دست خدا هستم. من ریسمان استوار خدایم و کلمه تقوایم و امین راز خدایم. من حجت عظمی و آیت کبری و نمونه والا و دروازهٔ پیامبر مصطفی هستم. من میزان اعمالم.



فصل اوّل

اندازه‌گیری و مدل سازی

آیا فکر می‌کنید میزان تحصیلات افرادی که با آنها تماس دارید در علاقهٔ شما به ادامهٔ تحصیل نقش تعیین‌کننده‌ای دارند؟ مسلماً بی‌تأثیر نیست. چگونه می‌توان تصور خوبی از بیشتر افرادی که روزانه با آنها برخورد می‌کنید در مورد تحصیلات عالی بدست آورد؟

آیا تاکنون از خود پرسیده‌اید روزانه با چه کسانی سر و کار دارید؟ این افراد چه سن و سالی دارند؟ شغل آن‌ها چیست و میزان تحصیلات آن‌ها چقدر است؟ برای چند روز متوالی فهرستی از این افراد تهیه کنید، اطلاعاتی دربارهٔ شغل، میزان تحصیلات و سن آن‌ها جمع‌آوری کنید.

قبل از آن که این فهرست را تهیه کنید چه اعدادی در ذهن شما مجسم شده است؟ پاسخ شما به این سؤال‌ها در ابتدا ممکن است حتی به‌صورت کمّی (عددی) نباشد، مثلاً ممکن است بگویید : روزانه با تعداد کمی از افراد سر و کار دارم، اکثر آن‌ها باسواد و جوان هستند. این پاسخ‌ها برای ترسیم دورنمای اولیّه از اطلاعات مورد نظر کفایت می‌کند، ولی به هیچ وجه از نظر مطالعات علمی سودمند نیستند. این اطلاعات باید با معیارهای مناسب **اندازه‌گیری** و به‌صورت اعداد و ارقام بیان شوند. برای انجام بررسی‌های آماری به اندازه‌گیری نیاز داریم.

اندازه‌گیری

فکر می‌کنید فردی کوتاه قد یا بلند قد هستید؟ هنگام پاسخ، قدتان را با چه کسانی مقایسه می‌کنید. اگر درچین و یا اتریش زندگی می‌کردید جواب‌های متفاوتی به این سؤال می‌دادید. اگر بخواهید در مورد قدتان دقیق‌تر صحبت کنید اوّلین قدم اندازه‌گیری آن است.

ما در بخش کشاورزی در تولید گندم در سال‌های اخیر پیشرفت زیادی داشته‌ایم.

طول قد شما چقدر است؟ قد را با کدام واحد اندازه می‌گیرید؟ متر؟ سانتی‌متر؟ میلی‌متر؟

اگر بتوانیم اطلاعات خود را به زبان اعداد و ارقام بیان کنیم، کار مفیدتری انجام داده‌ایم. چون اطلاعات کمّی (اطلاعاتی که به صورت اعداد و ارقام بیان شدند) قابل تفسیر نیستند و برای همه کس یک معنای واحد دارد، امّا اطلاعات کیفی (توصیفی) این‌طور نیستند ممکن است شنونده برداشت‌های متفاوتی از آن داشته باشد. تیترخبری بالا را درنظر بگیرید.

این جملهٔ خبری خوشحال کننده است ولی برای مسئولان که باید دربارهٔ واردات گندم تصمیم بگیرند کافی نیست. باید منظور از سال‌های اخیر معلوم باشد آیا منظور دو سال گذشته است و یا ۶ سال گذشته پیشرفت در تولید گندم چقدر بوده است؟ آیا به میزانی بوده است که ما را از واردات آن بی‌نیاز کند و یا اگر تا آن حد نبوده است به چه میزانی کمبود داریم. ما نمی‌توانیم به کشورهای صادرکننده بگوییم مقداری گندم به ما بفروشید. باید وزن و یا حتی نوع آن کاملاً مشخص باشد.

اولین اقدام در رسیدن به اطلاعات عددی اندازه‌گیری است. انتخاب معیار اندازه‌گیری برای بعضی از موضوعات از قبیل، قد، وزن، میزان تحصیلات و درجه حرارت آسان است و با وسایلی که می‌شناسیم می‌توانیم آن‌ها را اندازه‌گیری کنیم.

هرچند انتخاب معیار مناسب برای اندازه‌گیری ممکن است ابتدا آسان به‌نظر بیاید ولی برای برخی از موضوعات، کار آسانی نیست. به‌نظر می‌رسد به آسانی می‌توان قد، وزن، سن و میزان تحصیلات افرادی را که با آن‌ها سر و کار دارید، اندازه‌گیری کنید ولی میزان درآمد آن‌ها به این سادگی نیست. (امتحان کنید) مثلاً مردم زیاد مایل نیستند که دربارهٔ درآمد خود با دیگران صحبت کنند، بنابراین اگر شما بخواهید از میزان درآمد افراد اطلاعاتی کسب کنید، مشکلاتی خواهید داشت. در بعضی مواقع مردم **کم‌گویی** می‌کنند، یعنی درآمد خود را کم‌تر از آن چیزی که هست بیان می‌کنند و در بعضی مواقع **زیادگویی** می‌کنند. شاید به بعضی‌ها برخورد کنید که اصلاً نمی‌دانند میزان درآمد آن‌ها چقدر است ممکن است یک میزان تقریبی برای آن بیان کنند. در هر صورت شما سعی کنید درآمد افراد خانواده، برخی از افراد فامیل،



آشنایان و دوستان را معلوم کنید. در عمل خواهید دید که دست‌یابی به میزان درآمد به سادگی اندازه‌گیری قد نیست.

از این‌ها مشکل‌تر اندازه‌گیری موضوعاتی است که دسترسی به آن‌ها مشکلاتی دارد. برای مثال ممکن است شما دربارهٔ هوش همکلاسی‌های خود قضاوت کرده باشید. بعضی‌ها را باهوش و برخی دیگر را دارای هوش متوسط ارزیابی کرده باشید؛ ولی اگر بخواهید مقایسه دقیق‌تری انجام دهید، لازم است معیاری برای سطح برخورداری از هوش در اختیار داشته باشید تهیّه این معیارها و استاندارد کردن آن‌ها کار افراد متخصص است. آن‌ها به‌نحوی وسایل اندازه‌گیری را طراحی می‌کنند تا مطمئن شوند این وسیله **همان چیزی را اندازه می‌گیرد که مورد نظر آن‌هاست.**



فرض کنید هنوز وسیله‌ای برای اندازه‌گیری وزن درست نشده است و ما می‌خواهیم سنگینی افراد و یا اشیا را اندازه‌گیری کنیم. ممکن است این‌طور استدلال کنیم هر شیئی که جای بیش‌تری اشغال کند (حجم آن بیش‌تر باشد) سنگین‌تر است و لذا برای محاسبهٔ سنگینی اشیا، آن‌ها را در ظرف مدرج پر از آب قرار دهیم و به این ترتیب حجم آن را حساب کنیم. شاید این‌طور اندازه‌گیری‌کردن تا حدودی ما را به واقعیت نزدیک کند ولی می‌دانیم که دو مفهوم حجم و وزن دو مقولهٔ جدا از هم هستند. می‌دانیم می‌توان دو قطعه هم وزن از طلا چنان ساخت که یکی حجیم‌تر از دیگری به نظر برسد. بنابراین رسیدن به وزن از طریق حجم کار درستی نیست. در تعیین ظرفیت هواپیما و آسانسورها وزن اهمیت دارد نه حجم. دربارهٔ هوش اگر مواظب نباشیم شاید حافظه فرد و یا سایر قابلیت‌های دیگر ذهنی او اندازه‌گیری شود. در این کتاب، واردِ مبحث ساختن معیارها، برای اندازه‌گیری نمی‌شویم و به آنچه که تاکنون آشنا شده‌ایم و به‌کار بسته‌ایم قناعت می‌کنیم.

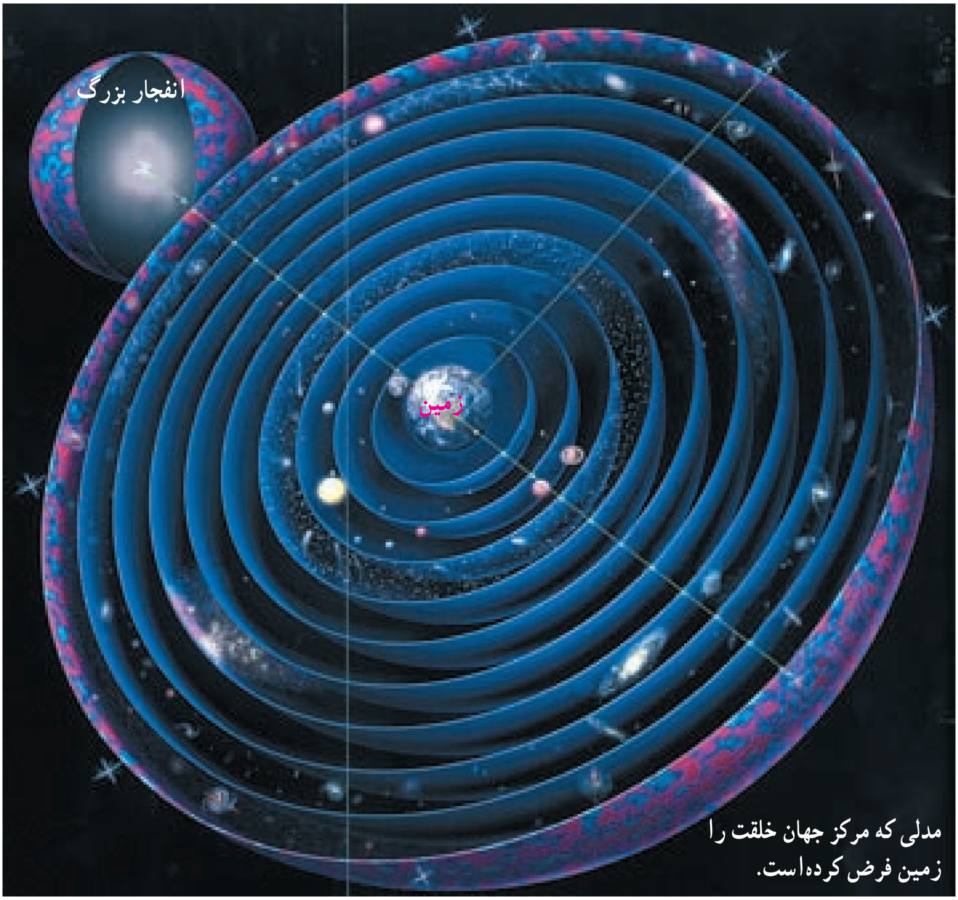
فعالیت

کاغذی در کلاس خود به دیوار نصب کنید و در مقابل آن بایستید، بالای سر خودتان را روی کاغذ علامت بزنید و اندازهٔ قد و نام خود را در کنار آن یادداشت کنید.

این ورقه می‌تواند به عنوان یادگاری از شما در این کلاس باقی‌بماند و سایرین هم در سال‌های آینده از آن تبعیت کنند. این اعداد ممکن است مبنایی برای مقایسه قد دانش‌آموزان در سال‌های مختلف تحصیل در این مدرسه باشد.

مدل‌سازی

کپلر ستاره‌شناس بزرگ در مورد خلقت جهان چنین گفته است : «**خداوند جهان را به زبان اعداد خلق کرده است**». شاید سخن او به این معنی است که هر آن چه که خداوند آفریده است به زبان ریاضی قابل توضیح و تفسیر است. مثلاً کاری که خود کپلر در طول ۲۳ سال کار مداوم انجام داد، بیان ریاضی علت و چگونگی حرکت سیارات در قالب سه قانون کپلر است. البته برای رسیدن به این سه قانون مجبور شد فرض‌هایی را درنظر بگیرد. مثلاً سیارات را به‌صورت نقطه درنظر گرفت. در حرکت، مسیرها و اندازه‌گیری‌ها از *تقریب‌های مناسب* استفاده کرد. همه ما زمین را به‌شکل کره در نظر می‌گیریم و حال آن که واقعیت امر این نیست. کره، ساده‌ترین و نزدیک‌ترین شئ ریاضی است که زمین به آن شباهت دارد. همین درنظر گرفتن شکل زمین به‌صورت کره نوعی مدل‌سازی ریاضی است. وقتی که کپلر قوانین حرکت سیارات را در قالب سه قانون خود با استفاده از نمادها و اصطلاحات و قوانین ریاضی بیان کرد در واقع حرکت سیارات را مدل‌سازی ریاضی کرد.



مدلی که مرکز جهان خلقت را زمین فرض کرده‌است.

شما اگر به بازی‌های کودکانه توجه کنید تمام آن‌ها به نوعی مدل‌های زندگی بزرگسالان است. وقتی کودکی عروسکی را در آغوش دارد همان رفتار و حرکاتی را دارد که مادران به هنگام در آغوش گرفتن فرزندان از خود بروز می‌دهند. وقتی کودکی اتومبیل کوچکی را روی زمین حرکت می‌دهد، ساده‌ترین شکلی است که در دنیای کودکانه می‌توان اتومبیل‌رانی را مدل‌سازی کرد. *زیبایی، خلاقیت و اهمیت این نوع بازی‌ها در این است که اولاً اسباب‌بازی‌ها حتی‌الامکان ساده باشند و ثانیاً کودکان بتوانند آنچه را که در ذهن آن‌ها می‌گذرد بهتر نشان دهند.* ما اگر بتوانیم با مفاهیم ریاضی و یا آمار پدیده‌ها را توضیح دهیم یک مدل‌سازی ریاضی و یا آماری کرده‌ایم.

بیان مسئله به زبان ریاضی را مدل‌سازی ریاضی می‌گوییم. هرچقدر مفاهیم ریاضی به‌کار برده شده ساده‌تر و ابتدایی‌تر و نتیجهٔ کار به پدیدهٔ موردنظر نزدیک‌تر باشد، مدل‌سازی، با ارزش‌تر است.

ما در این درس به نوعی خاص از مدل‌سازی ریاضی که براساس آمار و مفاهیم آن بنا شده‌است، اشاره خواهیم کرد. در واقع خواهید دید هر یک از مفاهیمی را که در آمار معرفی می‌کنیم ابزاری است برای مدلسازی پدیده‌های تحت مطالعه.

مثلاً قد افراد یک پدیدهٔ طبیعی است که بنابر ضرورت در مسائل مختلف در نظر گرفته می‌شود. وقتی شما قد یک فرد را اندازه می‌گیرید در واقع نزدیک‌ترین و ساده‌ترین مفهوم ریاضی به قد را که همان طول قد، که یک عدد است درنظر گرفته‌اید. پس از این اندازه‌گیری‌ها شما با افراد و سایر مشخصات آن‌ها کاری ندارید، فقط این اعداد هستند که از این پس موضوع مطالعهٔ شما خواهند بود.

وقتی یک طراح از یک چهره طراحی می‌کند در واقع مدلی از این چهره می‌سازد این مدل می‌تواند خیلی ساده باشد، مثلاً از چند خط راست تشکیل شده باشد مانند شکل الف.


الف


ب

اگر بتوانیم از خطوط منحنی و برخی از مهارت‌های طراحی هم استفاده کنیم مدل پیشرفته‌تری خواهیم داشت که به شکل واقعی نزدیک‌تر است مانند شکل ب. همین‌طور اگر ادامه دهیم و اجازه داشته باشیم از رنگ و سایر وسایل نقاشی استفاده کنیم به مدل کامل‌تری دست خواهیم یافت.

این تقریبات متوالی در اندازه‌گیری طول هم رخ می‌دهد. اگر فقط بخواهیم قد فردی را وجب کنیم (صرف‌نظر از اینکه اندازهٔ دست افراد متفاوت است)، مثلاً خواهیم گفت قد همکلاسی شما کمی بیش‌تر از هفت وجب خودش است. چون مقدار کمی بیش‌تر را نمی‌توانیم برحسب وجب بیان کنیم، خواهیم گفت قد این فرد هفت وجب است. و این یک مدل‌سازی ساده از قد است. ولی اگر قد همین فرد را با یک خط‌کش مدرج اندازه‌گیری کنید خواهید گفت قد این فرد ۱۵۴ سانتی‌متر است. مدل‌سازی اول خیلی ساده و ابتدایی بود و لذا کار دقیق نمی‌توان روی آن انجام داد. مثلاً شما لباسی را تصور کنید که برحسب وجب برای کسی دوخته شده است. امّا مدل‌سازی دوم دقیق است و لباسی که از این طریق دوخته می‌شود قابل قبول‌تر است. البته ما در اندازه‌گیری قد نتوانسته‌ایم کار را تمام کنیم، زیرا از واحدهای کوچک‌تر از سانتی‌متر صرف‌نظر کرده‌ایم. البته دو یا سه میلی‌متر در مقابل یک متر محسوس نیست. لذا صرف‌نظر کردن از آن‌ها لطمهٔ زیادی به مدل نخواهد زد.

ولی مواردی هست که به‌سادگی نمی‌توان از این جزئیات گذشت. شما وزن خود را برحسب کیلوگرم اندازه‌گیری می‌کنید. اگر بخواهید دقت زیاد داشته باشید از نیم‌کیلو هم استفاده خواهید کرد. از این رو وزن افراد معمولاً به‌صورت اعداد طبیعی (مثلاً ۵۳ کیلو، ۶۴ کیلو) و یا به‌صورت اعشار (مثلاً ۵۴/۵ کیلو و ۷۲/۵ کیلو) است. به ندرت از واحدهای گرم در اندازه‌گیری وزن بزرگسالان استفاده می‌شود. زیرا در مقابل چند ده کیلو، چند گرم به حساب نمی‌آید. امّا در مسابقات کشتی از این واحدهای جزیی یعنی گرم هم نمی‌توان گذشت. در شرایط مساوی کسی برنده است که وزن او ولو چند گرم، کم‌تر باشد. در مسابقات شنا به واحدهای دهم ثانیه هم متوسل می‌شوند. اگر روزی وسایل اندازه‌گیری دقیق‌تر شود شاید به واحد کوچک‌تر هم متوسل شوند ولی شاید هیچ‌گاه نتوانند مقدار دقیق زمانی را که یک شناگر ۵۰ متر را طی کرده است اندازه‌گیری کنند. *وسایل اندازه‌گیری را همیشه می‌توانند دقیق‌تر کنند ولی هیچ‌گاه دقت به آن اندازه‌ای نخواهد رسید که* ***خطای اندازه‌گیری*** *را صفر کند.*

خطای اندازه‌گیری همان تفاضل مقدار واقعی و مقدار اندازه‌گیری شده است. این خطا لزوماً از واحد اندازه‌گیری کمتر است.

چگونه می‌توانیم این پدیدهٔ عدم دسترسی به مقدار واقعی را مدل‌سازی کنیم؟ فرض کنید مدت زمانی که یک دونده فاصله ۱۰۰ متر را دویده است برابر ۷ ثانیه اندازه‌گیری شده باشد. می‌دانیم مدت‌زمان مذکور ممکن است دقیقاً برابر ۷ ثانیه نباشد، پس مرتکب خطایی که اندازهٔ آن از یک ثانیه کم‌تر است شده‌ایم، ولی واقعاً چقدر کم‌تر، نمی‌دانیم! از این رو اگر مدت زمانی که دونده ۱۰۰ متر را بر حسب ثانیه دویده است با متغیر T نمایش دهیم، می‌توان مدل زیر را برای آن پیشنهاد کرد :

$T = ۷ + E^*$ ثانیه

که در آن قدرمطلق E از یک ثانیه که واحد اندازه‌گیری ماست کم‌تر است، E ممکن است مثبت و یا منفی باشد.

به عنوان مثال دیگر فرض کنید وزن فردی برابر ۶۴/۵ کیلو گزارش شده باشد. چه مدلی ممکن است برای اندازه‌گیری بکار رفته باشد؟

فرض کنیم P وزن این فرد باشد، در این صورت

$P = ۶۴/۵ + E$

پس در این جا واحد اندازه‌گیری نیم‌کیلو در نظر گرفته شده است .

پس قدرمطلق E از ۰/۵ کیلو و یا ۵۰۰ گرم کم‌تر است.

اگر یک پدیده تابعی از چند متغیر باشد، خطای آن متغیرها چگونه در پدیدهٔ نهایی ظاهر می‌شود؟

مثلاً فرض کنید شعاع دایره‌ای به‌صورت $R = ۲ + E$ مدل‌سازی شده باشد. می‌خواهیم مدلی برای مساحت این دایره بنویسیم.

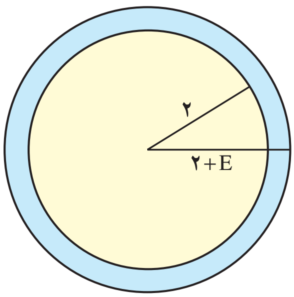
می‌دانیم مساحت دایره برابر $\pi R^۲$ است. بنابراین اگر فرض کنیم :

$S =$ مساحت دایره

آن‌گاه

$S = \pi(۲ + E)^۲$

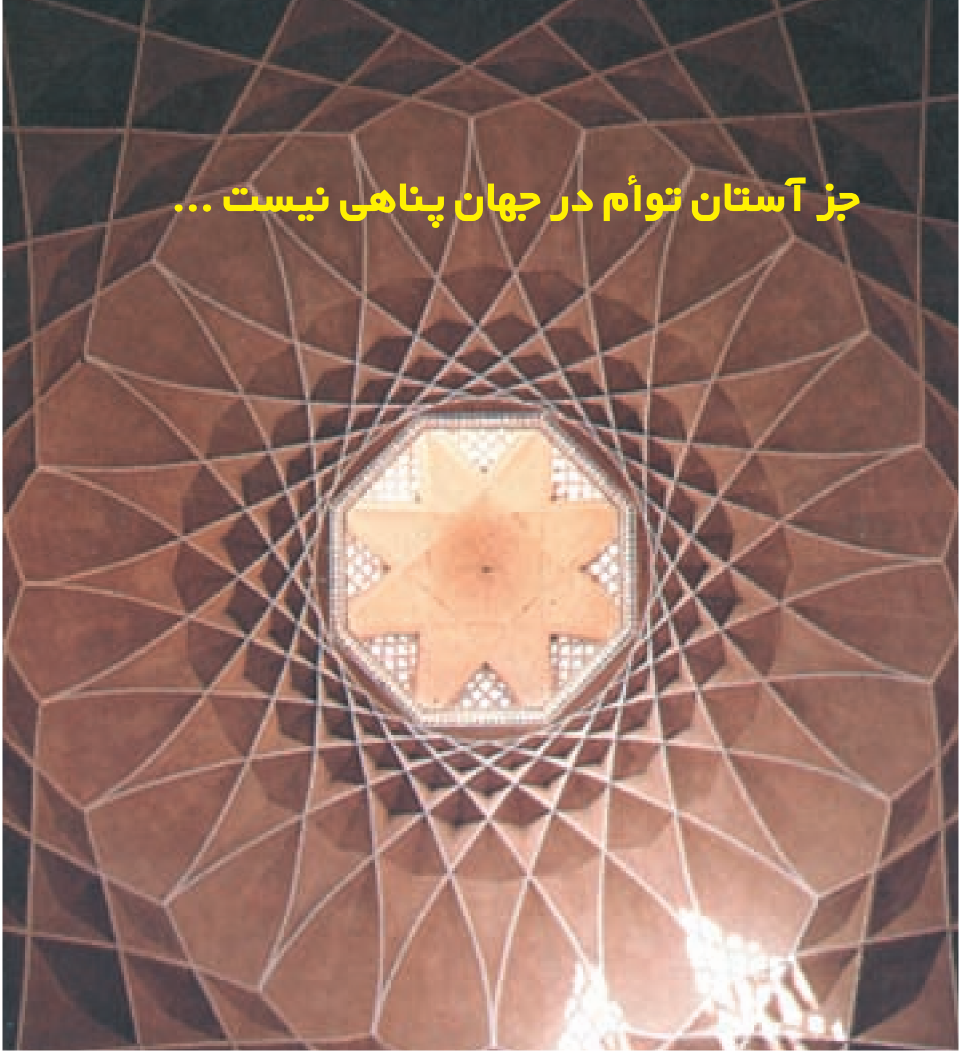
$= \pi(۴ + E^۲ + ۴E)$



می‌توان از جملاتی که شامل توان دوم یا بالاتر از خطاست صرف‌نظر کرد.

---

* E ابتدای ERROR به معنای خطا می‌باشد.



پس داریم :

$$S \approx 4\pi \pm 4\pi E$$

اگر $4\pi E$ را برابر E_1 فرض کنیم مدل مساحت دایره عبارت خواهد بود از :

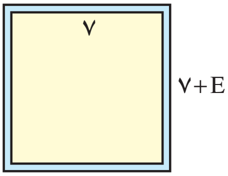
$$S \approx 4\pi \pm E_1$$

که 4π مساحت دایره‌ای به شعاع ۲ است.

می‌بینیم که خطای مساحت 4π برابر از خطای شعاع بزرگتر است. پس ممکن است از واحد سطح بزرگتر شود. برای جلوگیری از این اتفاق باید دقت اندازه‌گیری طول شعاع را بیشتر کنیم. می‌توان با اندازه‌گیری دقیق‌تر E را چنان کوچک کرد که E_1 از واحد سطح کمتر باشد.

حال به مثال دیگری می‌پردازیم. اگر ضلع مربعی به‌صورت $7+E$ باشد، مساحت آن با چه تقریبی در دسترس است؟

فرض کنید S برابر مساحت مربع باشد، در این صورت :



$$S=(۷+E)^۲=۴۹+۱۴E+E^۲$$

با صرف‌نظر کردن از جملات مرتبه بالا داریم :

$$S\approx ۴۹+۱۴E$$
$$=۴۹+E_۱$$

اگر بخواهیم ببینیم محیط مربع با چه تقریبی در دسترس است، این‌طور می‌نویسیم.

فرض کنید P محیط مربع باشد، در این صورت :

$$P=۴(۷+E)=۲۸+۴E$$
$$=۲۸+E_۲$$

ما در عمل محیط مربع را برابر ۲۸ خواهیم گرفت ولی می‌دانیم که محیط مربع با ۲۸ تفاوتی دارد. اگر قرار باشد این تفاوت را به حساب بیاوریم باید از برابری $P=۲۸+E_۲$ استفاده کنیم.

خطای ناشی از اندازه‌گیری اضلاع بیشتر در محیط مربع اثر می‌گذارد یا در مساحت آن؟

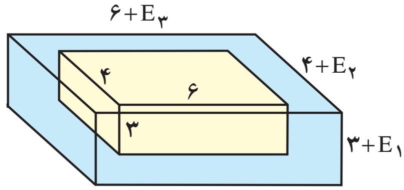
دربعضی موارد پدیده مورد بررسی متأثر از چند عامل است که هر یک با خطای متفاوتی اندازه‌گیری می‌شوند.

مثلاً اگر اضلاع یک مکعب مستطیل را به‌صورت زیر داشته باشیم :

ارتفاع $=L_۱=۳+E_۱$

عرض $=L_۲=۴+E_۲$

طول $=L_۳=۶+E_۳$



حجم این مکعب از چه مدلی پیروی می‌کند؟ فرض کنید حجم مکعب برابر V باشد، در این صورت :

$$V=L_۱L_۲L_۳$$
$$=(۳+E_۱)(۴+E_۲)(۶+E_۳)$$
$$\approx ۷۲+۲۴E_۱+۱۸E_۲+۱۲E_۳$$

در این مدل E_i ها از واحد کوچک‌ترند پس حاصل‌ضرب آن‌ها خیلی کوچک خواهد بود، پس می‌توانیم از آن‌هم صرف‌نظر کنیم. بنابراین یک مدل سادهٔ قابل قبولی به‌دست خواهد آمد. به پدیده‌های اطراف خود توجه کنید و ببینید تابع چه متغیرهایی هستند و سعی کنید اثر خطای اندازه‌گیری در هر یک از متغیرها را در کل پدیده بررسی کنید.

در دو مثال قبل دیدیم که چگونه خطای اندازه‌گیری در مساحت و محیط مربع و حجم مکعب مستطیل «منتشر» می‌شود. این روابط به ما کمک می‌کنند تا برای جلوگیری از خطاهای بزرگ در محصول نهایی محاسبات، از ابتدا پیش‌گیری‌های لازم را به‌عمل آوریم.

تمرین



۱ــ دماسنجی بردارید و با اندازه‌گیری دمای اتاق مدلی برای گرمای آن بنویسید. این مدل تا چه مدت اعتبار دارد؟ آیا از همین مدل برای ۸ ساعت دیگر نیز می‌توانید استفاده کنید؟ توضیح دهید.

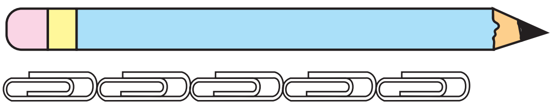
۲ــ طول مداد و قطر قاعدهٔ آن را برحسب میلی‌متر اندازه‌گیری کنید و مدلی برای حجم مداد بنویسید. (Eهایی که برای طول و قطر قاعده به‌کار می‌برید یکی نیستند، توضیح دهید.)

۳ــ طولی را برحسب سانتی‌متر اندازه‌گیری کرده‌اید و مدل آن را به‌صورت $L = ۲ + E$ نوشته‌اید. اگر این طول را برحسب میلی‌متر اندازه‌گیری کنید آیا مدل $L = ۲۰ + E$ برای آن مناسب است یا خیر؟ توضیح دهید.

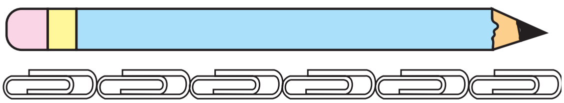
۴ــ مدلی برای وزن خودتان پیشنهاد کنید. با استفاده از ابزار اندازه‌گیری وزن، درستی آن را تحقیق کنید.

۵ــ مدلی برای اندازه‌گیری طول مداد مطابق با هر یک از شرایط زیر بنویسید.

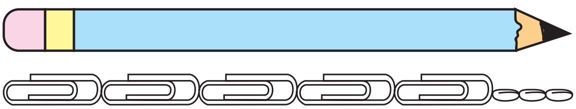
الف

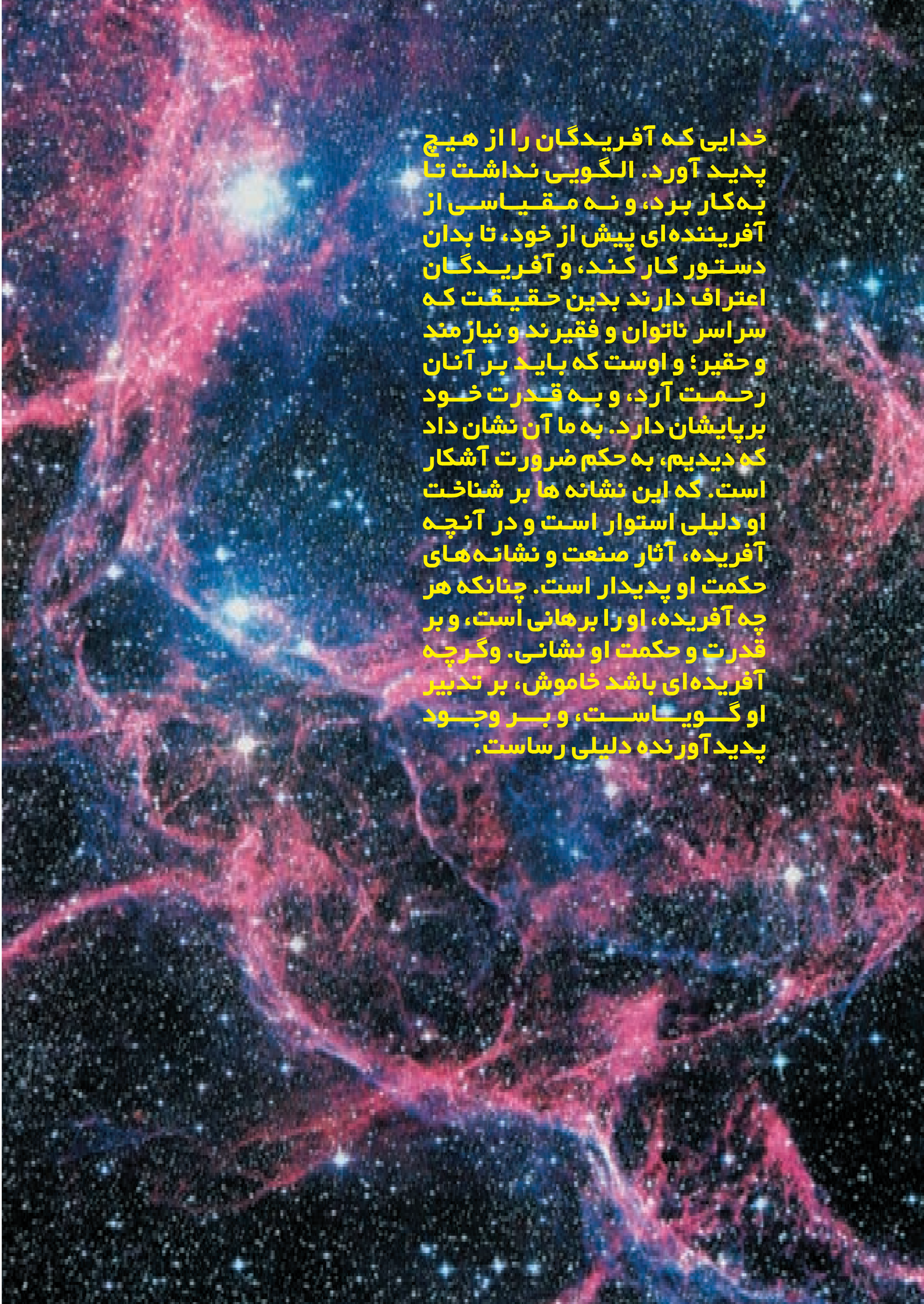


ب



ج





خدایی که آفریدگان را از هیچ پدید آورد. الگویی نداشت تا به‌کار برد، و نه مقیاسی از آفریننده‌ای پیش از خود، تا بدان دستور کار کند، و آفریدگان اعتراف دارند بدین حقیقت که سراسر ناتوان و فقیرند و نیازمند و حقیر؛ و اوست که باید بر آنان رحمت آرد، و به قدرت خود برپایشان دارد. به ما آن نشان داد که دیدیم، به حکم ضرورت آشکار است. که این نشانه ها بر شناخت او دلیلی استوار است و در آنچه آفریده، آثار صنعت و نشانه‌های حکمت او پدیدار است. چنانکه هر چه آفریده، او را برهانی است، و بر قدرت و حکمت او نشانی. وگرچه آفریده‌ای باشد خاموش، بر تدبیر او گویاست، و بر وجود پدیدآورنده دلیلی رساست.

جامعه و نمونه

چند ستاره در آسمان می‌توان یافت؟ آیا تا به‌حال در یک شب صاف که ماه در آسمان نیست به آسمان نگاه کرده‌اید؟ *شاید فکر کرده‌اید که تعداد ستاره‌ها آن‌قدر زیاد است که هرگز نمی‌توان آن‌ها را شمرد.* اما باید بدانید که تعداد ستارگانی که می‌توان بدون چشم مسلّح دید بیش از ۳۰۰۰ نیست. اگر با یک دوربین به آسمان نگاه کنید تعداد آن‌ها تا ۵۰۰۰۰ افزایش می‌یابد و اگر با یک تلسکوپ بزرگ به آسمان نظر کنید تعداد ستارگانی که در دیدرس شما هستند تا ۵۰۰,۰۰۰,۰۰۰ افزایش می‌یابند. این عدد آن‌قدر بزرگ است که اگر شما در تمام طول عمر هرشب تمام شب را به شمردن ستارگان بگذرانید، عمر شما کفاف نخواهد داد که این تعداد ستاره را بشمارید.

فکر می‌کنید ستاره‌شناسان چگونه به چنین عدد بزرگی رسیده‌اند؟ آنان این شمارش را با نمونه‌گیری انجام داده‌اند. می‌توانیم آسمان را به چندین بخش مساوی تقسیم کنیم و تعداد ستارگان را در برخی از این نواحی بشماریم. براساس این اعداد، می‌توان تعداد تقریبی ستاره‌ها را محاسبه کرد.

جامعه

در فصل قبل با مسائل مربوط به اندازه‌گیری و مدل‌سازی‌های آن به‌طور مقدماتی آشنا شدید. حال به موضوع قد افراد باز می‌گردیم. شما قبلاً قد بعضی از همکلاسی‌های خود را اندازه گرفته‌اید. اگر بخواهید راجع به قد افراد **مطالعه** کنید، چه اقداماتی انجام خواهید داد؟ خوب است که در یک کار علمی و دقیق، کلمات دارای معنای روشن و تعریف شده‌ای باشد. آیا می‌دانید منظور از مطالعه قد چیست؟ پاسخ این سؤال را به تدریج در طول این درس خواهید فهمید ولی مقدمتاً بیان می‌کنیم که منظور از مطالعهٔ قد افراد، آن است که بدانیم چند درصد افراد قد کوتاه و چند درصد دارای قد معمولی و چند درصد افراد قد بلند هستند و یا اصلاً می‌خواهیم بدانیم که چه افرادی را قدکوتاه و چه افرادی را قد بلند می‌گویند. اندازهٔ قد بیش‌تر مردم چقدر است؟ و سؤالاتی از این قبیل. بلافاصله ممکن است بر این سؤالات این اشکال را وارد کنید که سؤال‌های بالا برای سنین مختلف متفاوت

است. معیار قدکوتاهی برای نوجوانان با معیار قدکوتاهی برای جوانان و یا بالاتر تفاوت می‌کند. پس نمی‌توانیم بدون محدود کردن افراد، جواب مشخصی به‌سؤال‌های بالا بدهیم. مگر افراد را مشخص کنیم و برای آن‌ها شرایطی تعیین کنیم. مثلاً بگوییم منظور ما مطالعه قد نوجوانان پسر است (پسران بین ۱۴ــ۱۱ سالگی) با این شرایط **مجموعه‌ای** از افراد مشخص داریم که مطالعه قد آن‌ها معقول به‌نظر می‌رسد و اشکالات مطرح شده در بالا را ندارد. چنین مجموعه‌ای را یک **جامعهٔ آماری** می‌گوییم.

جامعه آماری مجموعه‌ای از افراد یا اشیا است که دربارهٔ اعضای آن می‌خواهیم موضوع و یا موضوعاتی را مطالعه کنیم.

مثال‌هایی از جامعه آماری و ویژگی‌های مورد بررسی آن‌ها می‌توانند به‌صورت زیر باشند :

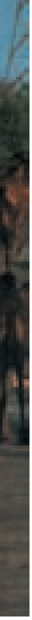
ــ محصولات کشاورزی استان مازندران و موضوع مورد مطالعه،انواع محصولات.

ــ واردات و صادرات کشور و موضوع مورد مطالعه، نوع کالاهای وارد شده یا صادرشده.

ــ مجموعهٔ دانش‌آموزانی که در سال ۱۳۷۹ در مقطع راهنمایی در شهر تهران ثبت نام کردند و موضوع مورد مطالعه، معدل سال قبل آن‌ها.

ــ مجموعهٔ پزشکان متخصص قلب و عروق و موضوع مورد مطالعه، تعداد بیمارانی که به علت ناراحتی‌های قلبی در مقطع معینی از سال به آن‌ها مراجعه کرده‌اند.







ــ مجموعهٔ کارگاه‌های خدمات فنی استان مرکزی و موضوع مورد مطالعه، تعداد کارگران شاغل در آن‌ها.

ــ مجموعهٔ افراد جویای کار در سه ماهه اول سال ۱۳۷۹ و موضوع مورد مطالعه، میزان تحصیلات آن‌ها.

ــ مجموعهٔ فارغ‌التحصیلان دانشگاه‌های ایران در سال ۱۳۷۸ و موضوع مورد مطالعه، طول دورهٔ تحصیل دانشگاهی آن‌ها.

ــ مجموعهٔ دبیران استان خراسان و موضوع مورد مطالعه، سابقهٔ تدریس آن‌ها.

ــ مجموعهٔ دبیران ریاضی استان خراسان و موضوع مورد مطالعه، سابقهٔ تدریس آن‌ها.

همانطوری که ملاحظه می‌کنید بعضی از جامعه‌های آماری خود زیر مجموعه‌ای از جامعه‌های آماری دیگر هستند. مثلاً مجموعهٔ دبیران ریاضی استان خراسان زیر مجموعه‌ای از دبیران این استان است.

نمونه

ما در مطالعه قد گفتیم که می‌خواهیم بدانیم چند درصد افراد با در نظر گرفتن معیار قد بلندی در جامعهٔ مورد بررسی، قد بلند هستند. یک راه حل ابتدایی آن است که *قد تمام افراد* را اندازه بگیریم و بعد ببینیم چند نفر قد بلند هستند و از این‌جا درصد مورد نظر را محاسبه کنیم.

مثلاً اگر جامعهٔ ما ۶۰۰ نفر عضو داشته باشد و ۶۰ نفر قد بلند باشند آن‌گاه ده درصد افراد این جامعه قد بلند هستند.





اگر تمام افراد جامعه را مورد مطالعه قرار دهیم می‌گوییم سرشماری کرده‌ایم.

معمولاً در سرشماری‌ها با مشکلاتی مواجه هستیم، مهم‌ترین آن‌ها عبارتند از :

ـ در دسترس نبودن تمام اعضای جامعه

ـ وقت‌گیر بودن دسترسی به تمام اعضای جامعه

ـ گران تمام شدن بررسی تمام اعضای جامعه

ـ از بین رفتن جامعه در برخی از مطالعات

این مشکلات سبب می‌شود تا سعی کنیم از راه میان‌بُر برویم و به‌جای آن که تمام اعضای جامعه را مورد مطالعه قرار دهیم بخشی از آن را که با دقت و مطالعهٔ لازم انتخاب شده است، بررسی کنیم. البته در این صورت بخشی از اطلاعات را از دست داده‌ایم ولی در مقابل، مشکلاتی را که در بالا به‌آن‌ها اشاره کردیم از پیش‌رو برداشته‌ایم. این بخش کوچک از جامعهٔ آماری را **نمونه** می‌گوییم.

نمونه زیر مجموعه‌ای از جامعهٔ آماری است.

فرض کنید بخواهید برخی از ویژگی‌های ماهی‌اوزن برون را در دریای خزر مطالعه کنید. در این صورت به تمام ماهی‌های دریای خزر دسترسی نخواهید داشت، لذا لازم است از طریق مطالعهٔ نمونه، این بررسی انجام شود.

در بعضی از کارخانه‌ها بخشی به‌نام بخش کنترل کیفی کالا وجود دارد که کالاها را از لحاظ مرغوبیت و یا سالم و معیوب بودن بررسی می‌کند. فرض کنید یک کارخانه تولید لامپ، بخواهد ولتاژی را که لامپ تولیدی آن کارخانه تحمل می‌کند بررسی کند. آزمایشی که برای تعیین سطح تحمل لامپ از آن استفاده می‌شود به‌گونه‌ای است که لامپ در این آزمایش می‌سوزد. در این مثال خاص سرشماری یعنی از بین بردن تمام محصولات کارخانه. پس برای جلوگیری از چنین زیانی مجبوریم از نمونه استفاده کنیم.

در ادبیات ما ضرب‌المثل «مشت نمونه خروار است» به‌خوبی موضوع جامعه و ضرورت استفاده از نمونه را نشان می‌دهد. در این جا مسأله عبارت است از خرید برنج برای مصرف یک سال خانواده. فرض کنید میزان مصرف ۱۰۰ کیلو باشد. قبل از خرید باید به قسمتی از کیفیت برنج موجود در انبار اطمینان حاصل کنیم. برای رسیدن به این هدف قسمتی از آن که در حد یک مشت برنج است آزمایش می‌شود و براساس آن تصمیم‌گیری می‌شود. در این شرایط برنج موجود در انبار جامعه و یک مشت



برنجی که انتخاب شده است، نمونه است.

ملاحظه می‌کنید که این نمونه نسبت به آن جامعه خیلی کوچک است. آیا نتیجه‌ای که از بررسی این نمونه کوچک حاصل می‌شود قابل اطمینان است؟ پاسخ به این سؤال مثبت است. زیرا فرض بر این است که برنج موجود در انبار همگی از یک نوع بوده و محصول یک دست است. اگر غیر از این باشد و برنج‌های موجود در انبار از انواع مختلف در گونی‌های متفاوت روی‌هم انبار شده باشند، این یک مشت نمی‌تواند معیاری برای قضاوت باشد.

عمل **نمونه‌گیری** مهم‌ترین بخش آمار را تشکیل می‌دهد. **یک نمونه** گروه کوچکی از اعضا است که به نحوی انتخاب شده‌اند که نمایان‌گر خصوصیات عده بزرگ‌تری که «**جامعه**» نام دارد، باشد.

برای آن که نمونه به‌درستی نمایانگر خصوصیت تمام جامعه باشد، باید به *اندازه کافی بزرگ* باشد. اگر تعداد اعضای نمونه خیلی کوچک باشد ممکن است اعضای آن نتوانند نمایندگان خوبی برای اعضای جامعه باشند. همچنین اعضای نمونه باید تصادفی انتخاب شوند. یعنی انتخاب آن‌ها نباید از قانون خاصی پیروی کند.

اندازهٔ جامعه و اندازهٔ نمونه: گفتیم جامعه یک مجموعه است که می‌خواهیم موضوعاتی را در آن مطالعه کنیم. این مجموعه اعضایی دارد. تعداد این اعضا در بررسی‌های آماری نقش مؤثری دارد. اگر جامعه در ارتباط با کاری که می‌خواهیم انجام دهیم کوچک باشد، بیشتر اوقات به‌صورت سرشماری مطالعه را انجام می‌دهیم ولی اگر بزرگ باشد همان‌طوری که گفتیم لازم است مطالعات از طریق نمونه‌گیری دنبال شود. از این رو تعداد اعضای جامعه را به‌عنوان پارامتری که باید مد نظر قرار داد مطرح می‌کنند. این تعداد را **اندازهٔ جامعه** می‌گویند.

تعداد اعضای جامعه را اندازه جامعه می‌گوییم.

بعضی از جامعه‌ها به سبب ویژگی خاص آن‌ها کوچک‌اند مثلاً مجموعهٔ افراد زال مو (افرادی که موهای آن‌ها از بدو تولد سفید است نه آنکه سفید شده باشد). این یک مجموعهٔ نسبتاً کوچک است و از جمله مواردی است که دسترسی به تمام آن‌ها امکان‌پذیر نیست.

مجموعهٔ ماهی‌های کفال دریای خزر، مجموعه‌ای است بزرگ و باز هم دسترسی به همهٔ آن‌ها غیر ممکن است.

مجموعهٔ دانش‌آموزان کشور، مجموعه‌ای بزرگ و ناهمگون است. زیرا آن‌ها را می‌توان

به‌دسته‌های پسر، دختر و یا از نظر مقطع تحصیلی به ابتدایی، راهنمایی و دبیرستان تفکیک کرد. حتی می‌توان از نظر محل جغرافیایی تحصیل نیز آن‌ها را تقسیم‌بندی کرد.

اندازهٔ جامعه در نمونه‌ای که می‌خواهیم انتخاب کنیم تأثیر دارد. اگر جامعه بزرگ باشد یعنی اگر اندازهٔ آن زیاد باشد، اغلب لازم است نمونه‌ای که از آن انتخاب می‌شود بزرگ باشد، از این رو تعداد اعضای نمونه هم یکی دیگر از پارامترهایی است که در یک بررسی آماری باید مورد توجه قرار گیرد. این تعداد را **اندازهٔ نمونه** می‌گوییم.

تعداد اعضای نمونه را اندازهٔ آن نمونه می‌گوییم.

هرچه جامعه بزرگ‌تر باشد، برای آن که نمونه بتواند بیان‌گر جامعه باشد، باید نمونه بزرگ‌تر باشد. همچنین اگر جامعه از گوناگونی بیش‌تری برخوردار باشد برای حصول اطمینان از حضور نماینده‌ای از این گوناگونی‌ها در نمونه، باز هم لازم است نمونه بزرگ باشد. اگر اندازهٔ نمونه مناسب نباشد نتیجه‌ای که از آن گرفته می‌شود معتبر نخواهد بود. مثلاً در شهری با جمعیت ۱۰ میلیون نفر اگر از ۱۰۰ نفر دربارهٔ یک موضوع اجتماعی سؤال شود، نتیجه‌ای که از آن حاصل می‌شود نمی‌تواند قابل تعمیم به جامعه باشد. بنابر ضرب‌المثل معروف «**با یک گل بهار نمی‌شود**» نمی‌توان همواره با شواهد کم حکم کلی کرد.

همان‌طوری که گفتیم اندازهٔ جامعه ممکن است خیلی بزرگ باشد، در بعضی مواقع حتی ممکن است نامتناهی باشد و یا آن‌قدر بزرگ باشد که ما آن را نامتناهی فرض کنیم، مطالعهٔ این قبیل جامعه‌ها مشکلات خاص خود را دارد و چون ما هنوز در ابتدای راه هستیم فقط **جامعه‌هایی را مورد مطالعه قرار می‌دهیم که اندازهٔ آن‌ها متناهی باشد.**



بحث کلاسی

در زیر چند نمونه‌گیری مطرح می‌شود. علت تناسب یا عدم تناسب این نمونه‌ها را در کلاس به بحث بگذارید. همچنین روشی برای نمونه‌گیری مناسب ارائه دهید.

۱ــ موضوع مورد مطالعه: بررسی سطح نمرهٔ ریاضیات در دورهٔ دبستان.

نمونه: دانش‌آموزان یکی از مناطق آموزش و پرورش که حرف اول اسم کوچکشان با «ذال» شروع می‌شود.

۲ــ موضوع مورد مطالعه: نظرسنجی درمورد اینکه کدام یک از زن و شوهر نقش مهم‌تری در ادارهٔ خانواده دارند.

نمونه: دبیران ریاضی دبیرستان‌های پسرانه شیراز.

۳ــ موضوع مورد مطالعه: مطالعه آماری تعداد فرزندان هر خانواده در شهر تهران.

نمونه: ساکنین یکی از مناطق تهران که به تصادف انتخاب می‌شوند.

۴ــ موضوع مورد مطالعه: رضایت مشتریان از محصولات یک کارخانه.

نمونه: خانواده‌های کارکنان آن کارخانه.

تمرین

۱ــ فرض کنید بخواهید میزان آلودگی هوا را در شهر خود بررسی کنید. جامعه را معرفی کرده و توضیح دهید که نمونه از این جامعه به چه صورتی انتخاب می‌شود.

۲ــ می‌خواهیم مدت زمانی را که طول می‌کشد تا افراد یک شهر از منزل به سر کارشان بروند، مطالعه کنیم، جامعه را معرفی کنید و دربارهٔ نمونه‌ای که باید انتخاب شود صحبت کنید.

۳ــ از یک تولیدکننده مرکبات می‌خواهید ۵۰ جعبه پرتقال بخرید. قبلاً می‌خواهید دربارهٔ اندازهٔ پرتقال‌ها مطمئن شوید. آیا بررسی پرتقال‌هایی که روی جعبه چیده شده است کفایت می‌کند؟

۴ــ از سه موضوع مورد مطالعه زیر دربارهٔ کلاسی که در آن درس می‌خوانید، کدام‌یک از تنوع بیش‌تری برخوردار است.

موضوع مورد مطالعه :

۱ــ معدل

۲ــ قد

۳ــ سن

در این تمرین، نمونهٔ مربوط به بررسی کدام یک از ویژگی‌ها می‌تواند کوچک باشد؟

۵ــ پنج خانوار از بین خانوارهای همکلاسی‌های خود انتخاب و تعداد افراد هر خانوار را تعیین کنید. با توجه به اطلاعات بالا به سؤال‌های زیر پاسخ دهید :

ــ بیش‌تر خانوارها چند نفره هستند؟

ــ خانوارهایی که بیش‌ترین تعداد عضو را دارند درنظر بگیرید. این خانوارها چند نفره هستند؟

قبل از این که اطلاعات را جمع‌آوری کنید حدس شما دربارهٔ پاسخ سؤال‌های بالا چیست؟ پاسخ خود را یادداشت کنید و آن را با نتیجه حاصل از اطلاعات بالا مقایسه کنید. آیا جمع‌آوری اطلاعات براساس حدس و گمان نتایج مستندی به‌دست می‌دهد؟

۶ــ سن پدر ۸ نفر از همکلاسی‌های خود را بپرسید. حال به سؤال‌های زیر پاسخ دهید :

ــ چند نفر از آن‌ها بین ۲۰ تا ۳۵ سال سن دارند؟

ــ چند نفر از آن‌ها بین ۳۵ تا ۴۵ سال سن دارند؟

ــ چند نفر از آن‌ها بیش از ۴۵ سال سن دارند؟

قبل از اینکه سن دبیران خود را بپرسید چه پاسخی برای سؤال‌های بالا داشتید، حال حدس خود را با این پاسخ‌ها مقایسه کنید و دلایل یکسانی و عدم یکسانی آن‌ها را جستجو کنید.

۷ــ ۱۵ نفر از همکلاسی‌های خود را انتخاب کنید و از آن‌ها دربارهٔ تعداد محصلین در خانوارشان سؤال کنید. این تعداد را یادداشت کنید و به سؤال‌های زیر پاسخ دهید.

ــ بیش‌ترین تعداد محصل در این خانوارها چقدر است؟

ــ بیش‌تر این خانوارها چند محصل دارند؟

۸ــ «خانواده‌ها کوچک‌تر شده‌اند. سال‌ها قبل متوسط تعداد فرزندان خانواده‌ها ۵ بوده است ولی اکنون ۲ است.» فرض کنید از شما خواسته شده تا یک نمونهٔ ۱۰۰ تایی از خانواده‌ها را درنظر گرفته و درستی خبر فوق را تحقیق کنید. در این مطالعه : جامعه و نمونه را تعریف کنید.

۹ــ ترس از بی‌کاری برای بخشی از مردم وجود دارد. بررسی وضعیت شغلی کارمندان شاغل در یک شرکت خصوصی، نشان داد از هر ۵ نفر، ۱ نفر نگران از دست دادن شغل خود است. در این بررسی، جامعه و نمونه را مشخص کنید.

۱۰ــ «کارمندان بیمارستان‌ها ترجیح می‌دهند به‌جای نوبت روز، در نوبت شب کار کنند.»

نظر ۳۰ پرستار شاغل در نوبت شب چند بیمارستان راجع به کار کردن در نوبت شب پرسیده شد. پاسخ آن‌ها به قرار زیر است :

۱۱ نفر نوبت روز را ترجیح دادند. ۱۶ نفر نوبت شب را ترجیح دادند. ۲ نفر بی‌تفاوت بودند.

از این بررسی می‌توان دید که پرستاران ترجیح می‌دهند نوبت شب کار کنند. در این بررسی و نتیجه‌گیری چه ایرادهایی وجود دارد؟



نمونه تصادفی

نمونه تصادفی ساده: کلاس خود را به عنوان جامعه آماری در نظر بگیرید. اندازهٔ آن چقدر است؟ یک نمونه با اندازه ۸ از آن انتخاب کنید. چگونه این نمونه را انتخاب کردید؟ یکی از راه‌هایی که ممکن است به‌کار برده باشید، آن است که ۸ نفر اول را به‌ترتیب حروف الفبا از روی دفتر کلاس انتخاب کرده باشید. ممکن است ۸ نفری را که اطراف شما در کلاس می‌نشینند، انتخاب کرده باشید. بعضی مواقع ممکن است روش انتخاب نمونه مشکل‌ساز باشد. ابتدا یادآوری می‌کنیم که هدف ما از نمونه و مطالعه آن شناخت جامعه است. پس نمونه باید به‌قسمی انتخاب شود که بتواند «بیانگر» جامعه باشد. با این مقدمه فرض کنید بخواهید دربارهٔ قد همکلاسی‌های خود مطالعه کنید. حال اگر افراد تیم بسکتبال کلاس را به‌عنوان نمونه انتخاب کرده باشید، نمونه خوبی دربارهٔ قد همکلاسی‌های خود ارائه نکرده‌اید. آیا می‌توانید بگویید چرا؟ دلایل خود را توضیح دهید.

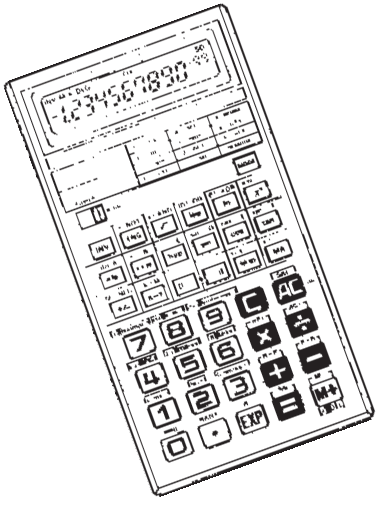
باید روش انتخاب نمونه به‌گونه‌ای باشد که :

ــ امکان انتخاب هر فرد به‌عنوان عضوی از نمونه امکان‌پذیر باشد.

ــ قبل از انتخاب نمونه، نتوانیم با اطمینان بیش‌تر دربارهٔ حضور و یا عدم حضور عده‌ای در نمونه قضاوت کنیم. به عبارت دیگر، هر فرد برای شرکت در نمونه همان‌قدر سهم داشته باشد که دیگران دارند.

نمونه‌ای که ویژگی‌های بالا را داشته باشد **نمونه تصادفی ساده** می‌گوییم و برای دستیابی به‌آن‌ها روش‌هایی ذکر شده است که یکی از آن‌ها استفاده از اعداد تصادفی است که ذیلاً توضیح می‌دهیم.

اعداد تصادفی: این عنوان شاید بنابر آن‌چه که تا به حال خوانده‌ایم، درست نباشد و درواقع همچنین است چون ما عدد تصادفی نداریم. آن چه که در این عنوان نهفته است این است که ما عددی را از بین مجموعه‌ای از اعداد با اتکا به روش‌های تصادفی انتخاب کنیم. منظور از روش‌های تصادفی، یعنی روشی که این اطمینان را به ما بدهد که سهم انتخاب عددی از سایرین بیش‌تر و یا کم‌تر نیست. مثلاً اگر از فردی بخواهید که عددی طبیعی از بین اعداد ۱۰ تا ۱۵ انتخاب کند، به‌ندرت ممکن است عدد انتخاب شده ۱۳ باشد. زیرا اعتقادات مردم دربارهٔ این عدد به گونه‌ای است که کم‌تر ممکن است به آن توجه کنند و معمولاً در انتخاب، اعداد ابتدایی و انتهایی سهم کم‌تری برای انتخاب شدن دارند. پس انتخاب تصادفی به روشی که اراده و سابقه ذهنی انسان در آن دخیل باشد، عملی نیست. از این رو با برنامه‌نویسی‌هایی که در ماشین‌های حساب و کامپیوترها شده است، می‌توان به اعدادی که خارج از کنترل، سلیقه و سابقه ذهنی انسان‌ها باشد دسترسی پیدا کرد. ما ذیلاً به روش دست‌یابی به اعداد تصادفی در ماشین حساب می‌پردازیم.



این روش در اکثر ماشین حساب‌ها یکسان به‌کار می‌رود. پس از روشن کردن ماشین حساب، دکمه inv (در برخی، دکمه SHIFT) را فشار دهید. سپس دکمه RAN# را فشار می‌دهیم. پس از انجام این اعمال، عددی مثلاً به‌صورت ۰/۳۵۷ در صفحه ماشین ظاهر می‌شود. حال اگر مراحل فوق را تکرار کنیم اعداد تصادفی دیگری به دست خواهد آمد. اعداد تصادفی که توسط ماشین حساب تولید می‌شوند همواره غیر منفی و کوچکتر از ۱ است.

فرض کنید عدد تصادفی انتخاب شده ۰/۲۷۳ باشد و اندازهٔ جامعه‌ای که می‌خواهیم از آن نمونه را انتخاب کنیم ۱۵۰ باشد. در این صورت:

$$۰/۲۷۳ \times ۱۵۰ = ۴۰/۹۵$$

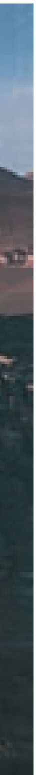
قسمت اعشاری عدد را حذف کرده و یک واحد به آن اضافه کنید. ۴۱، یکی از اعدادی است که از بین اعداد از ۱ تا ۱۵۰ به تصادف انتخاب شده است.

انتخاب نمونه: به مسئله انتخاب ۸ نفر از بین همکلاسی‌های خود باز می‌گردیم. تعداد افراد

کلاس را می‌دانید، مثلاً ۳۵ نفر. به این افراد می‌توانید خودتان یک شماره از ۱ تا ۳۵ اختصاص دهید و یا آن‌که همان شماره‌ای را که در دفتر کلاس دارند، درنظر بگیرید. حال با استفاده از ماشین حساب یک عدد تصادفی تولید کنید. مثلاً ۰/۱۲۴ در این صورت با ضرب این عدد در ۳۵ خواهیم داشت ۴/۳۴ = ۳۵×۰/۱۲۴. با حذف قسمت اعشاری و اضافه کردن ۱ به ۴ فرد شماره ۵ به‌عنوان عضوی از نمونه انتخاب خواهد شد. این عمل را ۸ بار تکرار می‌کنیم. جدول زیر اعداد تصادفی که توسط کامپیوتر تولید شده است را نشان می‌دهد.

جدول ۱

شماره	۱	۲	۳	۴	۵	۶	۷	۸
عدد تصادفی	۰/۴۹۶	۰/۳۹۹	۰/۹۴۳	۰/۱۰۸	۰/۵۹۳	۰/۱۹۹	۰/۹۷۷	۰/۲۹۸
۳۵ ×	۱۷/۳۶	۱۳/۹۶۵	۳۳/۰۰۵	۳/۷۸	۲۰/۷۵۵	۶/۹۶۵	۳۴/۱۹۵	۱۰/۴۳
شمارهٔ انتخاب شده	۱۸	۱۴	۳۴	۴	۲۱	۷	۳۵	۱۱



با ضرب کردن در ۳۵ اعداد تصادفی نامنفی و کوچکتر از ۳۵ به‌دست آمده است.

بنابراین، نفرات به شماره‌های ۱۴،۱۸ و ... ۱۱ اعضای نمونهٔ شما را تشکیل خواهند داد.

تذکر: ممکن است در تولید اعداد تصادفی این حالت پیش بیاید که دو شمارهٔ یکسان به‌دست آید. در این صورت از یکی از این اعداد صرف‌نظر کرده و عمل را یک بار بیش‌تر تکرار می‌کنیم تا در نهایت ۸ عدد متمایز داشته باشیم.

آن چه که از نمونه موردنظر ماست خود افراد یا اشیای انتخاب شده نیستند. قبلاً در تعریف جامعه گفتیم جامعه مجموعه اشیایی است که در یک یا چند موضوع مشترک باشند. نمونه که زیرمجموعه‌ای از جامعه است نیز این ویژگی را دارد. حال که نمونه را انتخاب کرده‌ایم، باید سراغ موضوعات مورد توجه برویم. مثلاً اگر بخواهیم قد افراد کلاس را مطالعه کنیم، باید به اندازه‌گیری قد آن‌ها بپردازیم.

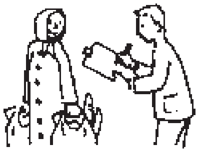
روش‌های جمع‌آوری داده‌ها

شما کلاس خود را به‌عنوان جامعه درنظر گرفتید ۸ نفر از همکلاسی‌های خود را به‌عنوان یک نمونهٔ ۸ تایی با استفاده از اعداد تصادفی انتخاب کردید. موضوع مورد مطالعه شما قد همکلاسی‌ها بود. حال قد همکلاسی‌های خود را اندازه بگیرید. حاصل کار ۸ عدد خواهد بود. این اعداد را که حاصل اندازه‌گیری اعضای نمونه هستند، **داده** می‌گوییم.

نتایج حاصل از اندازه‌گیری و یا بررسی نمونه را *داده* می‌گوییم.

داده‌ها به چند طریق جمع‌آوری می‌شوند :

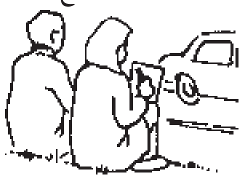
۱ــ استفاده از داده‌های از پیش تهیه شده



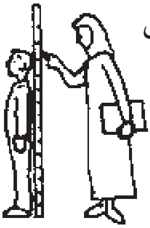
۲ــ از طریق پرسش : مستقیماً از اشخاص (شفاهی، مصاحبه)

: پرسش‌نامه کتبی

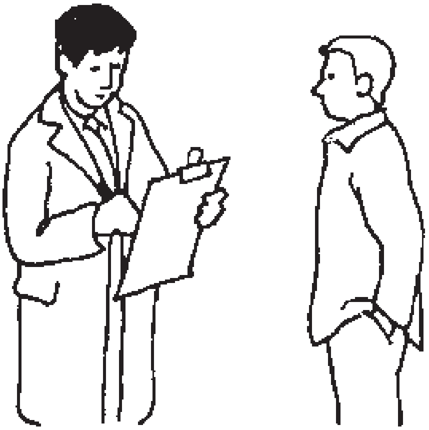
۳ــ از طریق مشاهده و ثبت وقایع



۴ــ از طریق انجام آزمایش

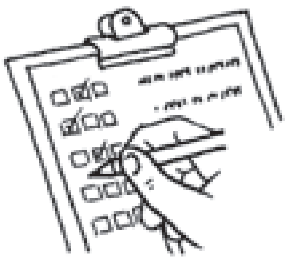






طراحی پرسش‌نامه: پرسیدن سؤال یکی از راه‌های متداول جمع‌آوری داده است که به دو صورت مصاحبه یا پرسش‌نامه کتبی می‌باشد. برای آن که مطمئن شویم از تمام افرادی که مورد بررسی هستند، سؤالاتی به‌طور یکسان پرسیده می‌شود، پرسش‌نامه‌ای به‌صورت کتبی تهیه می‌شود. از طرفی دیگر در مصاحبه اگر فرد مورد بررسی سؤال را به خوبی متوجه نشود مصاحبه‌گر می‌تواند توضیحات بیش‌تری بدهد.

ــ قبل از آن که هرگونه سؤالی نوشته شود، محتوای پرسش‌نامه باید سازمان‌دهی شود.



ــ هدف بررسی باید در نظر گرفته شود. چه می‌خواهید بدانید و چرا؟

ــ فهرستی از عناوینی که باید راجع به آن‌ها اطلاعات جمع‌آوری کنید، تهیه کنید.

ــ از جمع‌آوری داده‌ها و اطلاعات اضافی که مورد نیاز نمی‌باشند خودداری کنید.

ــ از سؤالات ساده و کاملاً واضح استفاده کنید. استفاده از عباراتی چون بزرگ، کوچک که معیار مشخصی ندارند خودداری کنید تا از سؤالات چند برداشت نشود.

سعی کنید از سؤالاتی که پاسخ آن‌ها «تا حد امکان» یک کلمه‌ای یا اعداد هستند استفاده شود.

ــ در مورد سؤالاتی که پاسخ‌دهنده ممکن است نخواهد جواب دقیق بدهد مانند سن، درآمد و ... از پاسخ‌های پیش آماده شده چون : کمتر از ۱۶ سال

۱۶ الی ۲۵ سال

۲۶ الی ۳۵ سال

۳۶ الی ۴۵ سال

۴۶ الی ۵۴ سال

بالاتر از ۵۴ سال استفاده کنید.

ــ از سؤالات هدایت‌کننده استفاده نکنید. به‌طور مثال اگر می‌خواهید برنامهٔ جدید مدرسه را بررسی کنید، هرگز نپرسید : نظرتان را راجع به برنامهٔ کامل شدهٔ مدرسه بیان کنید. عبارت «برنامهٔ کامل

شده» به‌طور غیرمستقیم در خواننده نظر مثبت القا می‌کند.

ــ سعی کنید از سؤالات با پاسخ‌های چندگزینه‌ای استفاده کنید.

ــ اگر جویای نظر یا عقیده راجع به موضوع یا محصولی هستید، پاسخ‌ها را به‌صورت گزینه‌های کمّی یا کیفی سطح‌بندی شده انتخاب کنید. مانند : بسیار خوب ☐

خوب ☐

متوسط ☐

ضعیف ☐

بسیار ضعیف ☐

ــ دستورالعمل کامل و واضح پاسخ‌گویی به پرسش‌نامه را ضمیمه کنید.

ــ در پایان پرسش‌نامه از پاسخ‌دهندگان تشکر کنید.

برای به انجام رساندن یک بررسی آماری باید پاسخ سؤالهای زیر را بدانید:

۱ــ می‌خواهید چه چیزی پیدا کنید؟ چرا؟

۲ــ چه داده‌هایی را باید جمع‌آوری کنید؟ چگونه تصمیم می‌گیرید؟

۳ــ کدام راه، بهترین راه جمع‌آوری داده است؟ چرا؟ آیا به‌وسیله و ابزار خاصی نیاز دارید؟

۴ــ به چه مقدار داده نیاز دارید؟ چگونه تصمیم می‌گیرید؟

۵ــ داده‌های شما باید چقدر دقت داشته باشند؟ چرا؟ آیا می‌توانید داده‌هایتان را آزمایش کنید؟

۶ــ داده‌هایتان را چگونه ثبت می‌کنید؟ چرا این روش را انتخاب کردید؟

۷ــ داده‌ها را چگونه ارائه می‌کنید؟ چرا؟

۸ــ آیا داده‌ها از الگوی خاصی پیروی می‌کنند؟ داده‌ها بیان‌گر چه می‌باشند؟

۹ــ چه نتیجه‌گیری یا پیش‌گویی می‌توانید از داده‌ها بکنید؟ آیا می‌توانید نتیجه یا پیش‌گویی خود را آزمایش کنید؟

۱۰ــ آیا نتایج، با آنچه انتظار داشتید تطبیق دارد؟

۱۱ــ چگونه نتایج کار را ارائه می‌کنید؟ برای چه کسانی ارائه می‌کنید؟

۱۲ــ با توجه به نتایج به‌دست آمده، آیا می‌توانید سؤالات دیگری را نیز مورد بررسی قرار دهید؟

به‌نظر شما در مرحلهٔ جمع‌آوری داده به کدامیک از سؤالات فوق باید پاسخ داده شود؟

فعالیت

ما معمولاً از یک مشکل زمانی آگاهی پیدا می‌کنیم که یک واقعهٔ غیرقابل قبول یا غیرقابل انتظار گزارش داده شود و اگر این مشکل به اندازه کافی بزرگ باشد در روزنامه‌ها و یا دیگر وسایل ارتباط جمعی منعکس می‌شوند. به مثال زیر توجه کنید :

خبر زیر در یکی از روزنامه‌ها آمده است :

مادری با فرزندش مجبور شدند ۲۰ دقیقه صبر کنند تا از عرض یک اتوبان بگذرند.

اهالی اطراف یکی از بزرگراه‌ها از این که مجبورند زمان زیادی را منتظر بمانند تا با ایمنی از عرض اتوبان بگذرند خسته شده‌اند. این مشکل هفتهٔ گذشته به شهرداری منطقه پس از این که مادری مجبور شد ۲۰ دقیقه صبر کند تا با فرزندش از عرض اتوبان عبور کند، گزارش داده شد ......

فرض کنید، شما مسئول بررسی این مشکل هستید. ابتدا باید بدانید مشکل تا چه حد جدی است. سپس باید اطلاعاتی چون چند نفر در روز نیاز به عبور از عرض اتوبان در این محل دارند؟، به‌طور متوسط چند دقیقه باید صبر کنند تا به‌طور اَمن از عرض اتوبان بگذرند؟، آیا همیشه عبور از عرض اتوبان مشکل است یا گاه‌گاه؟، ... جمع‌آوری کنید.

پرسش‌نامه‌ای طرح کنید تا بتوانید این اطلاعات را جمع‌آوری کنید. (آیا به جز پرسش‌نامه می‌توان از روش دیگری استفاده کرد؟)

جامعه در این بررسی چیست؟ و نمونه خود را چگونه انتخاب می‌کنید.

چه راه حل‌هایی برای رفع این مشکل پیشنهاد می‌کنید؟

تمرین

۱ــ یک نمونه ۸ تایی از بین اعداد ۱ تا ۶۲۵ انتخاب کنید.

۲ــ یک نمونه ۶ تایی از بین اعداد طبیعی ۲۳ تا ۷۸ انتخاب کنید.

۳ــ مجموعهٔ افراد فامیل درجه اول و درجه دوم خود را بنویسید و آن‌ها را شماره‌گذاری کنید. سپس یک نمونه ۴ تایی از این جامعه انتخاب کنید.

۴ــ موضوع‌های زیر را در نظر بگیرید. دربارهٔ بهترین روش جمع‌آوری داده برای آزمایش این فرضیه‌ها تصمیم بگیرید :

ـ بیش‌تر مردم فکر می‌کنند «حداکثر سرعت در اتوبان‌ها باید تعیین شود».
ـ آبی، رنگ مورد علاقهٔ بیش‌تر مردم برای ماشین است.
ـ در زمان مطالعه گوش دادن به موسیقی کلاسیک به یادگیری کمک می‌کند.
ـ بیش‌تر تصادفات اتومبیل‌ها را رانندگان با سن کم‌تر از ۲۵ سال موجب می‌شوند.
ـ رژیم گرفتن، موجب کاهش هوش می‌شود.

۵ـ دو موضوع بررسی برای هر یک از روش‌های جمع‌آوری داده بیان کنید.
ـ از طریق مصاحبه
ـ از طریق مشاهده
ـ از طریق مطالعهٔ داده‌های از پیش تهیه شده
ـ از طریق آزمایش

پروژه

روی موضوعات زیر فکر کنید. برای هر موضوع که انتخاب می‌کنید، فهرستی از فرضیات در نظر بگیرید. یک پرسش‌نامه یا آزمایش طراحی کنید. نمونه‌ای برای بررسی انتخاب کنید و پرسش‌نامه را توزیع کنید. گزارشی از چگونگی و روش جمع‌آوری داده‌های خود تهیه کنید.
ـ اثر آلودگی روی کیفیت زندگی
ـ در دسترس بودن پارک‌های بازی کودکان اطرافِ محل زندگی‌تان
ـ زمانی را که کودکان در گروه‌های سنی مختلف صرف تماشا کردن تلویزیون می‌کنند.
ـ آزمون‌هایی که در بعدازظهر برگزار می‌شوند نتیجه بهتری از آزمون‌هایی که صبح‌ها برگزار می‌شوند دارند.
ـ گل‌ها هر چه نور خورشید بیش‌تری دریافت کنند، سریع‌تر رشد می‌کنند.
ـ هر چه وزن بیش‌تر باشد، سریع‌تر می‌توان وزن کم کرد.
ـ گوش دادن به موسیقی کلاسیک، یادگیری را افزایش می‌دهد.
ـ دانش‌آموزانی که نمره ریاضی بالاتر از ۱۸ دارند، نمرهٔ ادبیات بالای ۱۸ کسب می‌کنند.





و اگر به دیدهٔ دل بنگری بدانچه از بهشت برایت ستایند، دل برکنی از آنچه در دنیاست، هرچند بدیع و زیباست؛ از خواهش‌های نفسانی، و خوشی‌های زندگانی و منظره‌های آراسته. و فکر تو درمی‌ماند در آهنگ و آوای شاخه‌ها و برگ‌های درختان که در کنار نهرهای بهشتی است، و در آویختن خوشه‌های لؤلؤ آبدار بر شاخه‌های آن درختانی که ریشه‌های آن در پشته‌های مُشک نهان است و رُستن این میوه‌های گوناگون، در غلافها و پوشش‌های درون. شاخه‌ها بی‌رنجی خم گردد و چنانکه چیننده آن خواهد در دسترس او بود. خدای به رحمتِ خود، ما و شما را از آنان گرداند که به دل کوشد تا خود را به منزل‌های نیکوکاران رساند.






فصل سوم

متغیرهای تصادفی

مقدمه

تلویزیون یکی از رسانه‌های دیداری است که نقش بزرگی در فرهنگ هر جامعه‌ای دارد. تأثیر تلویزیون در آموزش مدرسه‌ای یک مسئلهٔ مهم است که می‌توان به کمک آمار آن را بررسی کرد. معلم بررسی نقش تلویزیون در تحصیل دانش‌آموزان را به‌عنوان پروژهٔ گروهی پیشنهاد کرد. یک گروه از دانش‌آموزان روی این مسئله کار کردند. با هم روند کار علی یکی از اعضای گروه را مرور می‌کنیم :

علی برنامه‌های شبکه‌های سیمای جمهوری اسلامی ایران را به ۱۰ دسته تقسیم نمود که عبارتند از : فرهنگی و اجتماعی، ورزشی و تفریحات سالم، معارف اسلامی، اطلاعات عمومی، هنر و ادب، سیاسی، اقتصادی، علم و فن، ارزش‌های دفاع مقدس، و در آخر برنامه‌های تاریخی. او با یک نمونه‌گیری از برنامه‌های سه ماه متوالی، تعداد ساعاتی را که هر شبکه به هر یک از موضوعات بالا اختصاص داده است، تخمین زد. علی برای این‌که مطمئن شود تعداد اعضای نمونه او کافی بوده است، بار دیگر نمونه‌ای به‌همان تعداد اوّل گرفت و نتایج به‌دست آمده را با نتایج اوّلیه مقایسه کرد. چون این نتایج اختلاف چندانی نداشتند نتیجه گرفت که نمونه او به تعداد کافی عضو داشته که بتواند نمایان‌گر تمام برنامه‌های تلویزیون باشد. اما این سؤال برای او پیش آمد که چرا نتایج دو نمونه‌گیری او «دقیقاً» منطبق بر هم نیستند؟ ما می‌دانیم که اطلاعات هر عضو نمونه، متغیری است که به عمل تصادفی نمونه‌گیری بستگی دارد. به همین دلیل موضوع مورد مطالعه را متغیر تصادفی می‌گوییم. چون مقدار آن پس از یک آزمایش تصادفی به‌دست می‌آید.

آن‌چه بین نتایج دو نمونه‌گیری تصادفی مناسب مشترک است الگوهای آماری است که در هر دو نمونه برقرار هستند. نتایج دو نمونه‌گیری تصادفی هرگز دقیقاً مساوی نیستند. حال دقت کنید به‌این‌که آیا علی واقعاً موضوع مورد سؤال را بررسی کرده است؟ بهتر بود علی پیش از جمع‌آوری داده‌ها با هم‌گروهی‌های خود مشورت می‌کرد!

در زیر میزان تولید برنامه در شبکه‌های سیمای جمهوری اسلامی ایران به نقل از گزارش‌های سالانه مرکز آمار آورده شده‌است.

جدول ۱ ــ میزان تولید برنامه در شبکه‌های سیمای جمهوری اسلامی ایران (ساعت)

سال و برنامه	جمع	شبکه اول	شبکه دوم	شبکه سوم	شبکه چهارم	شبکه پنجم (تهران)
۱۳۷۶	۱۱۸۷۶	۳۱۶۰	۲۵۹۶	۳۵۳۹	۱۱۴۱	۱۳۴۴
فرهنگی و اجتماعی	۴۸۰۵	۱۸۳۱	۹۷۵	۱۱۲۳	۲۸۳	۵۹۴
ورزش و تفریحات سالم....	۲۷۸۴	۲۱۰	۴۶۳	۲۰۷۶	۰	۳۶
معارف اسلامی............	۹۶۰	۲۵۷	۲۹۹	۱۰۴	۲۳۷	۶۳
اطلاعات عمومی..........	۱۲۰۹	۲۰۱	۴۲۶	۱۸۱	۲۰۹	۱۹۳
هنر و ادب................	۵۱۰	۰	۱۰۶	۱۸	۱۰۲	۲۸۵
سیاسی....................	۴۵۰	۱۷۲	۱۶۷	۲۹	۵۵	۲۸
اقتصادی..................	۴۳۷	۱۵۹	۵۴	۰	۱۳۷	۸۷
علم و فن..................	۳۱۸	۴۲	۷۰	۰	۱۹۷	۹
ارزش‌های دفاع مقدس...	۳۰۷	۲۲۹	۲۷	۸	۲۰	۲۳
تاریخی....................	۹۶	۵۹	۹	۰	۱	۲۶

متغیر تصادفی

در فصل قبل با جامعهٔ آماری و نمونه آشنا شدید. دیدیم برای آن‌که نمونه بتواند بیان‌گر جامعه باشد، باید به دور از نظر شخصی و سلیقه‌های فردی انتخاب شود، و برای این منظور از اعداد تصادفی استفاده کردیم و گفتیم حال که نمونه به‌دست آمده است، باید موضوع و یا موضوعات موردنظر را بررسی کنیم. این موضوع و یا موضوعات را **متغیر تصادفی** می‌نامیم.

تا به‌حال با متغیرهای تصادفی زیادی سر و کار داشته‌اید بدون آن‌که آن‌ها را با این نام بشناسید. قد دانش‌آموزان یک کلاس، سابقه کار دبیران ریاضی مدرسه شما، تعداد بیماران مراجعه‌کننده به یک درمانگاه، میزان درآمد افراد، تعداد کلماتی که در هر بیت از غزل‌های حافظ به‌کار رفته است، متغیرهای تصادفی هستند. حتی رنگ چشم افراد و یا گروه خونی متغیرهای تصادفی هستند.

مثال‌های دیگری از متغیرهای تصادفی عبارتند از :

ــ درجه حرارت هوا در ساعت ۱۰ صبح روزهای آبان ماه سال جاری.

ــ تعداد افراد خانواده همکلاسی‌های شما.

ـ مرحله زندگی افرادی که برای گردش به یک پارک محلی آمده‌اند (مراحل زندگی عبارت است از : نوزادی، کودکی، نوجوانی، جوانی، میان‌سالی و پیری).

ـ گروه خونی همکلاسی‌های شما.

برخی از متغیرها قابل اندازه‌گیری‌اند، مانند قد، وزن، درجهٔ حرارت، درآمد خانوار، میزان بارندگی در سال در یک منطقه، میزان آلودگی هوا، شدت زلزله و غیره. آن‌چه که در این متغیرها مهم است آن است که اگرچه ما نمی‌دانیم که مقدار حاصل از اندازه‌گیری چه‌قدر خواهد شد ولی می‌دانیم عدد حاصل در مجموعه‌ای خواهد بود. مثلاً قد نوجوانان می‌تواند از ۱۲۰ سانتی‌متر شروع و به ۲۰۰ ختم شود. درجه حرارت در منطقه‌ای در فصل معینی باز دامنه معلومی دارد.

برخی دیگر از این متغیرها قابل شمارشند. مانند تعداد افراد خانواده، تعداد زنبورهای یک کندو، تعداد غایبین کلاس، تعداد تصادف‌های رانندگی در یک روز و در یک شهر، تعداد اتومبیل‌هایی که در یک روز به علت تخلف رانندگی جریمه شده‌اند و تعداد تخم‌مرغ‌هایی که فردی در طول یک ماه مصرف می‌کند. این قبیل متغیرها مانند حالت قبل این ویژگی را دارند که اگرچه تعداد مورد نظر از قبل برای ما معلوم نیست ولی می‌توانیم محدوده‌ای برای آن مشخص کنیم.

برخی از دیگر متغیرها قابل اندازه‌گیری نیستند مانند گروه خونی، بلکه این متغیرها به نوع و یا دسته خاصی تعلق دارند، در این‌جا هم اگرچه از قبل نمی‌توانیم بگوییم که گروه خونی افراد انتخاب شده کدام نوع خواهد بود ولی مجموعهٔ انواع گروه خونی برای ما معلوم است(A، B، AB و O) مثال‌های دیگری از انواع متغیرهای غیرقابل اندازه‌گیری عبارتند از : کشت که ممکن است یکی از دو نوع دیم و یا آبی باشد، جنسیت افراد که ممکن است یکی از دو نوع زن و یا مرد باشد، کیفیت کالای تولید شده به‌وسیله کارخانه‌ای که ممکن است سالم و یا معیوب باشد، فردی از یک شهر ممکن است باسواد و یا بی‌سواد باشد و یا همچنین این فرد ممکن است شغل دولتی داشته باشد و یا شغل آزاد.





همان‌طوری‌که دیدیم متغیرها عموماً از دو نوع بودند یکی متغیرهایی که اندازه‌گیری می‌شدند و عدد به آن‌ها نسبت داده می‌شد که این متغیرها را **متغیرهای کمّی** می‌گوییم.

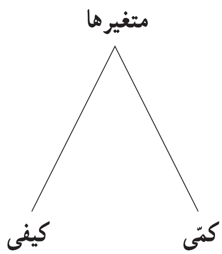
مثال‌هایی از متغیرهای کمّی عبارتند از : وزن، فاصله، طول، جمعیت، خسارت مالی ناشی از تصادفات، تعداد کسانی که در یک کلاس معدل بالای ۱۵ دارند، تعداد درس‌هایی که شما بالاتر از ۱۵ دارید. از ویژگی‌های یک متغیر کمّی آن است که می‌توانیم افراد را از نظر این متغیر مقایسه کنیم.

متغیرهای کمّی متغیرهایی هستند که قابل اندازه‌گیری‌اند.

نوع دیگری از متغیرها، آن‌هایی هستند که اندازه‌گیری نمی‌شوند، فقط نوع آن‌ها تعیین می‌شود، مانند گروه خونی، کشت و مراحل زندگی یک فرد که این نوع متغیرها را **متغیرهای کیفی** می‌گوییم.

متغیرهای کیفی متغیرهایی هستند که قابل اندازه‌گیری نیستند.

بنابراین ما متغیرها را از نظر این‌که قابل اندازه‌گیری باشند و یا نباشند به‌صورت زیر دسته‌بندی کردیم.



تمرین

۱ــ چهار متغیر تصادفی نام ببرید و نوع هر یک را از نظر کمّی و کیفی بودن تعیین کنید.

۲ــ فهرستی از متغیرهای تصادفی که روزانه با آن‌ها سر و کار دارید تهیه کنید. آیا در این فهرست حجم غذای مصرفی را منظور کرده‌اید، آیا میزان کالری دریافتی از این غذا را در نظر گرفته‌اید، میزان پروتئین را چطور؟ مدتی را که صرف تماشای تلویزیون می‌کنید ثبت کرده‌اید، مدتی را که به ورزش می‌پردازید، مدتی را که به مطالعهٔ درس‌های فردا می‌پردازید، مدتی را که مطالعه آزاد می‌کنید و...، به‌نظر می‌رسد این فهرست بی‌انتها می‌باشد. **ما به‌وسیله متغیرهای تصادفی زیادی احاطه شده‌ایم.**

انواع متغیرهای تصادفی: در بخش قبلی متغیرهای تصادفی معرفی شدند و آن‌ها را به دو دستهٔ عمدهٔ کمّی و کیفی تقسیم کردیم. اگر مثال‌های مربوط به متغیرهای کمّی را درنظر بگیرید ملاحظه خواهید کرد که بعضی از این متغیرها می‌توانند هر مقداری بین دو عدد را اختیار کنند، مثلاً وزن. اگر وزن فردی در کلاس شما ۶۰ کیلو و وزن یکی دیگر ۶۶ کیلو باشد، هر عددی را که بین ۶۰ و ۶۶ در نظر بگیرید (اعم از طبیعی و یا غیرطبیعی) دلیلی نداریم که وزن فردی نتواند برابر آن مقدار باشد. یعنی وزن می‌تواند به‌طور پیوسته تمام مقادیر واقع در بازه‌ای را بگیرد. این قبیل متغیرها را **متغیرهای کمّی پیوسته** می‌گوییم.

متغیر پیوسته، یک متغیر کمّی است که اگر دو مقدار a و b را بتواند اختیار کند هر مقدار بین آن‌ها را نیز بتواند اختیار کند.

در مقابل، متغیرهای کمّی دیگری هستند که ویژگی بالا را ندارند، مثلاً تعداد غایبین کلاس که ممکن است امروز ۵ نفر و دیروز ۳ نفر باشد، در این کلاس ممکن است روزی تعداد غایبین برابر ۴ نفر باشد ولی نمی‌تواند این تعداد برابر ۴/۵ شود. این قبیل متغیرها را **متغیرهای کمّی گسسته** می‌گویند.

به متغیر کمّی که پیوسته نباشد، گسسته گوییم.

معمولاً متغیرهای گسسته از نوع تعداد هستند مانند تعداد تصادفات، تعداد کالاهای سالم، تعداد روزهای بارانی در یک شهر و غیره، ولی ممکن است متغیرهایی نیز مثال زد که از نوع تعداد نباشند. مثلاً در ساختمان‌سازی، معمولاً ساختمان را در چند طبقه بنا می‌کنند ولی گاهی اوقات ممکن است یک طبقه را کامل نسازند که در این صورت به آن نیم‌طبقه گفته می‌شود. از این‌رو اصطلاح «ساختمانی در دو و نیم طبقه و یا سه و نیم طبقه» رایج است، در این صورت اگر متغیر تصادفی شما تعداد طبقات باشد، این متغیر مقادیر ۱ و ۱/۵ و ۲ و ۲/۵ و غیره را می‌تواند اختیار کند.

ــ بیش‌تر میوه‌ها بر اساس وزن که یک متغیر پیوسته است خرید و فروش می‌شوند. میوه‌هایی را مثال بزنید که بر اساس وزن خرید و فروش می‌شوند.

متغیرهای کیفی را نیز می‌توان به دو دسته تقسیم کرد. قبلاً در ویژگی‌های متغیرهای کمّی گفتیم که می‌توانیم افراد را از نظر متغیرهای کمّی مقایسه کنیم. مثلاً بگوییم فردی از فرد دیگری بلند قدتر است، سنگین وزن‌تر است، این کلاس بیش‌تر از کلاس دیگر غایب دارد. امروز آلودگی هوا کم‌تر از دیروز است و غیره.

اما چنین مقایسه‌ای را به‌وسیله متغیرهای کیفی نمی‌توانیم انجام دهیم، دو فرد از نظر گروه

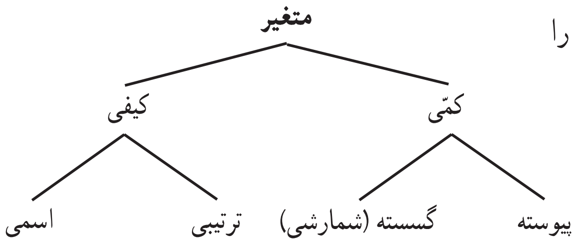
خونی مقایسه نمی‌شوند فقط می‌توان گفت که آیا این دو از نظر گروه خونی یکی هستند یا خیر. علت این امر در این است که در اعداد رابطهٔ بزرگ‌تری و کوچک‌تری تعریف شده است اما مثلاً در گروه خونی چنین رابطه‌ای وجود ندارد. اما بعضی از متغیرهای کیفی از نوعی ترتیب طبیعی برخوردارند مثلاً مراحل زندگی که در آن مسلماً نوزادی قبل از کودکی، کودکی قبل از نوجوانی و الی آخر است و یا مراحل تحصیل که در آن‌ها دبستان قبل از راهنمایی و راهنمایی قبل از دبیرستان و غیره است. این قبیل متغیرها را **متغیرهای کیفی ترتیبی** می‌گوییم.

متغیرهایی کیفی‌ای که در آن‌ها نوعی ترتیب طبیعی وجود دارد متغیرهای کیفی ترتیبی می‌گوییم.

متغیرهایی کیفی نیز وجود دارند که در آن‌ها این ترتیب ملاحظه نمی‌شود، مانند گروه خونی، RH خون. این قبیل متغیرها را **متغیر کیفی اسمی** می‌گوییم.

متغیر کیفی که ترتیبی نباشد متغیر کیفی اسمی می‌گوییم.

حال می‌توانیم تقسیم‌بندی متغیرها را به‌صورت روبرو تکمیل کنیم.



ــ غیر از مثال‌های ارائه شده در متن، برای هر یک از انواع متغیرها مثالی ارائه کنید.

ــ چند متغیر کیفی در توصیف میوه‌ها نام ببرید.





تمرین

۱ــ کدام‌یک از متغیرهای زیر گسسته و کدام‌یک پیوسته‌اند؟

ــ تعداد مکالمات تلفنی یک اداره در یک روز.

ــ طول مکالمات تلفنی یک اداره.

ــ تعداد نامه‌های یک صندوق.

ــ وزن نامه‌های موجود در یک صندوق.

ــ زمانی که یک بیمار در اتاق انتظار مطب یک پزشک منتظر است.

ــ تعداد بیماران مراجعه‌کننده به یک پزشک در طول روز.

۲ــ کدام‌یک از متغیرهای زیر کمّی و کدام‌یک کیفی هستند؟

ــ نوع تلفن مورد استفاده شهروندان (تلفن سیار و یا تلفن ثابت).

ــ جنسیت افراد یک شهر (زن، مرد).

ــ میزان تحصیلات افراد یک شهر (دیپلم، کارشناسی، کارشناسی ارشد، دکترا).

ــ وضع سواد (باسواد، بی‌سواد).

ــ وضعیت مسکن (صاحب مسکن، بدون مسکن ملکی).

ــ میزان اجاره پرداختی به‌وسیله شهروندان.

ــ میزان پرداخت مالیات سالانه ساختمان‌های مسکونی.

ــ میزان آلودگی هوا.

ــ میزان بارندگی در یک شهر در طول سال.

۳ــ نوع هر یک از متغیرهای زیر را مشخص کنید :

ــ رنگ اتومبیل‌های موجود در یک نمایشگاه اتومبیل.

ــ مقاومت یک ترانزیستور.

ــ گنجایش آب یک تانکر.

ــ تعداد شکایات رسیده شده به یک پاسگاه پلیس.

ــ درآمد دانشجویان شاغل به کار.

ــ وضعیت تأهل کارمندان یک شرکت.

ــ سن دانشجویان شرکت‌کننده در یک دورهٔ هنری.

۴ ــ کدام‌یک از نمونه‌گیری‌های زیر، یک نمونه‌گیری تصادفی است؟ در هر یک جامعه و نمونه را مشخص کنید.

ــ با تمام پلیس‌های یک پاسگاه پلیس برای پیدا کردن نظر پلیس‌های این پاسگاه راجع به تخلفات مصاحبه شد.

ــ با بچه‌هایی که وارد یک پارک بازی می‌شدند پنج در میان مصاحبه شد تا وسیلهٔ بازی مورد علاقهٔ کودکان را مورد بررسی قرار دهیم.

ــ برای بررسی pH شامپوهای تولیدی یک کارخانه، شامپوها را صد در میان مورد آزمایش قرار دادیم.

۵ ــ برای تحقیقات و بررسی‌های زیر جامعه را مشخص کرده و روش نمونه‌برداری پیشنهاد کنید.

ــ پیدا کردن درصد بیکاران استان مازندران.

ــ مردان کدام شامپو را بیش‌تر ترجیح می‌دهند.

۶ ــ می‌خواهیم مدت زمانی را که دانش‌آموزان کلاس شما در طول یک هفته صرف مطالعه کتاب‌های غیردرسی می‌کنند بررسی کنیم.

ــ در این بررسی جامعه را مشخص کنید.

ــ یک روش نمونه‌گیری برای انتخاب نمونه معرفی کنید.

ــ از چه روشی برای جمع‌آوری اطلاعات استفاده می‌کنید.

ــ برای این بررسی یک نمونه مشخص کنید.

ــ اندازه این جامعه چه‌قدر است؟

ــ اندازهٔ نمونهٔ مورد بررسی چه‌قدر است؟ آیا این تعداد برای بررسی مورد نظر مناسب است؟ در صورت مناسب نبودن، روش نمونه‌برداری و نمونه خود را اصلاح کنید.

ــ متغیر مورد مطالعه در این مسئله چیست؟ توضیح دهید.

ــ متغیر مورد مطالعه از چه نوعی است؟

پروژه

با افزایش جمعیت و با توجه به محدود بودن منابع آب، ضرورت صرفه‌جویی در مصرف آب احساس می‌شود. برای مطالعه در این مسئله مدلی را معرفی کنید.

لقمان و مرد پیاده



روزی لقمان در کنار چشمه‌ای نشسته بود. مردی که از آنجا می‌گذشت از لقمان پرسید : چند ساعت دیگر به ده بعدی خواهم رسید؟ لقمان گفت : راه برو. آن مرد پنداشت که لقمان نشنیده است. دوباره سؤال کرد : مگر نشنیدی؟ پرسیدم چند ساعت دیگر به ده بعدی خواهم رسید؟ لقمان گفت : راه برو. آن مرد پنداشت که لقمان دیوانه است و رفتن را پیشه کرد. زمانی که چند قدمی راه رفته بود، لقمان به بانگ بلند گفت : ای مرد، یک ساعت دیگر بدان ده خواهی رسید. مرد گفت : چرا اول نگفتی؟ لقمان گفت : چون راه رفتن تو را ندیده بودم، نمی‌دانستم تند می‌روی یا کند. حال که دیدم دانستم که تو یک ساعت دیگر به ده خواهی رسید.

در این داستان ساده و قدیمی تمام اصول آماری رعایت شده است.

جامعه؛ مجموعهٔ افرادی است که از محل زندگی لقمان به ده می‌روند.

متغیر تصادفی؛ مدت زمانی که طول می‌کشد تا فردی به ده برسد.

نمونه؛ این مرد نمونه‌ای از جامعهٔ افرادی است که به ده می‌روند و چند قدمی که مرد راه می‌رود نمونه‌ای است از جامعه قدمهای فرد تا شهر تا لقمان بتواند سرعت شخص را تخمین بزند.

نکته ظریف این داستان این است که لقمان فقط می‌گوید، راه برو و توضیح دیگری نمی‌دهد. لقمان نمی‌گوید که می‌خواهم راه رفتن تو را ببینم تا از روی آن بگویم چه مدت طول می‌کشد تا به ده برسی. زیرا لقمان فکر می‌کند این اطلاع ممکن است در راه رفتن آن مرد اثر بگذارد و در نتیجه سرعتی که لقمان تخمین می‌زند، سرعت واقعی راه رفتن آن فرد نباشد و در نتیجه زمانی را که تخمین خواهد زد، مدت زمان دقیقی نباشد.



سرلوحهٔ دین شناختن اوست؛ و درست شناختن او، باور داشتن او؛ و درست باورداشتن او؛ یگانه انگاشتن او؛ و یگانه انگاشتن او، او را بسزا اطاعت نمودن، و بسزا اطاعت نمودن او، صفتها را از او زدودن، چه هر صفتی گواه است که موصوف دو تاست و هر موصوف نشان دهد که از صفت جداست، پس هرکه پاک خدای را با صفتی همراه داند او را با قرینی پیوسته، و آن که با قرینش پیوندد، دوتایش دانسته؛ و آن که دو تایش خواند، جزء جزءاش داند؛ و آن که او را جزءجزء داند، او را نداند؛ و آن که او را نداند در جهتش نشاند؛ و آن که در جهتش نشاند، محدودش انگارد؛ و آن که محدودش انگارد، معدودش شمارد.

فصل چهارم

دسته‌بندی داده‌ها و جدول فراوانی

تاکنون با جامعه، نمونه، متغیر تصادفی و داده‌ها آشنا شدیم. پس از جمع‌آوری داده‌ها تعدادی عدد، مثلاً ۵۰ عدد در اختیار داریم. این اعداد داده‌های خام هستند. نگه‌داری این داده‌ها و استفاده از آن‌ها اگرچه جزیی‌ترین اطلاعات را دربر دارد ولی در مسایل بزرگ کنار آمدن با این همه اطلاعات جزیی غیرممکن و گیج‌کننده است. باید از بین این اطلاعات ریز و درشت یک طرح کلی از نمونه و نهایتاً از جامعه‌ای که این اعداد و ارقام از آن آمده‌اند، به‌دست آورد.

برای آن که موضوع مورد بحث ما روشن شود مثال زیر را درنظر بگیرید. شما در شهر و مکانی که زندگی می‌کنید اطراف خود را به‌خوبی می‌شناسید. نام کوچه‌ها، خیابان‌ها، مراکز خرید و حتی در بعضی از فروشگاه‌ها، مکان اجناس را نیز می‌دانید. حال اگر بخواهید محله خود را به کسی که با آن آشنایی ندارد، معرفی کنید ارائه این اطلاعات ریز نه مورد توجه است و نه آن که طرحی مناسب از محل زندگی شما به او می‌دهد. آن چه که با ارائه این اطلاعات نصیب شنونده خواهد شد، تصوری درهم از محل زندگی شماست. شما می‌توانید با ارائه شاخص‌های مناسبی از یک محله، چگونگی محل زندگی خود را معرفی کنید. مثلاً بگویید :



ــ مراکز خرید «متعدد» و در «دسترس» است.
ــ قیمت اجناس در آن «مناسب» است.
ــ مراکز تفریحی و فرهنگی «خوبی» دارد.
ــ میزان آلودگی هوا در آن «کم» است.
ــ سر و صدای شهر در آن‌جا «زیاد» نیست.

این اطلاعات به شنونده احساسی از مکان زندگی شما می‌دهد و همین‌ها کفایت می‌کند. سایر اطلاعات جز آن که ایجاد اخلال در سیستم اطلاع‌رسانی شما بکنند، فایده و تأثیر دیگری ندارد. در شاخص‌های معرفی شدهٔ بالا برخی از عبارات با نماد «   » مشخص شده‌اند. شما ممکن است ایراد بگیرید که اگر قرار است اظهارنظر علمی باشد باید با عدد و رقم همراه باشد. چگونه است که در

هیچ‌یک از این اظهارنظرها عدد و رقم دیده نمی‌شود. حق با شماست. ما باید با مطالعات قبلی تعریف کرده باشیم که مراکز فروش در «دسترس» هستند یعنی چه؟ میزان آلودگی هوا «کم» است یعنی چه؟ این عبارات بایستی به کمک اعداد و ارقام تعریف شوند. به‌خاطر سپردن اعداد و ارقام مخصوصاً زمانی که تعداد آن‌ها زیاد باشد، کار آسانی نیست. عبارات توصیفی را بهتر می‌توان به ذهن سپرد. از این‌رو برای انتقال سریع اطلاعات و نه برای منظورهای تخصصی، روش توصیفی مناسب است ولی این بیان توصیفی نمی‌تواند مبنایی برای تصمیم‌گیری‌های جدی شود. درهرحال ما ناگزیر هستیم برخی از اطلاعات را کنار بگذاریم. برخی دیگر را یک کاسه کنیم تا بتوانیم جامعهٔ مورد مطالعه را معرفی کنیم. این معرفی شبیه آن است که با چرخ بال از بالای شهر به آن نگاه کنیم. از این طریق نه وارد جزئیات شده‌ایم و نه آن که ویژگی‌های اصلی شهر را نادیده گرفته‌ایم.

جدول فراوانی

معمولاً داده‌ها در قالب یک جدول داده می‌شوند. این داده‌ها به علت طبیعت ظاهریِ نامنظم آن‌ها گویای مطلبی دربارهٔ جامعه نیستند. برای آن‌که بتوان به آن‌ها نظم بهتری داد جدول‌های مناسبی تنظیم می‌شوند. برای آن‌که موضوع بحث را بهتر ادامه دهیم، به مثال‌های زیر توجه کنید :

مثال: برای مطالعهٔ تعداد افراد خانوارهای یک شهرک، یک نمونه ۴۰تایی از خانوارهای این شهرک را انتخاب کرده و از آن‌ها تعداد افراد خانوار را پرسیده‌ایم. جدول زیر به‌دست آمده است :

۱ ,۲ ,۳ ,۵ ,۵ ,۲ ,۱ ,۲ ,۴ ,۳

۵ ,۳ ,۴ ,۱ ,۳ ,۳ ,۴ ,۴ ,۲ ,۶

۱ ,۶ ,۲ ,۴ ,۵ ,۵ ,۴ ,۲ ,۳ ,۷

۸ ,۷ ,۳ ,۶ ,۵ ,۴ ,۲ ,۴ ,۳ ,۳

برای آن‌که بتوانیم با نظم بخشیدن به داده‌های موجود اطلاعات بهتری کسب کنیم آن‌ها را در جدولی به‌صورت زیر خلاصه می‌کنیم :

تعداد افراد خانوار	۱	۲	۳	۴	۵	۶	۷	۸	جمع
تعداد خانوارها	۴	۷	۹	۸	۶	۳	۲	۱	۴۰

این جدول نشان می‌دهد که در نمونه، ۴ خانوارِ ۱ نفره، ۸ خانوارِ ۴ نفره و فقط ۱ خانوارِ ۸ نفره بوده است.

چون نمونه را به‌گونه‌ای انتخاب کرده‌ایم که بیان‌گر جامعه باشد می‌توانیم بگوییم در این شهرک خانوارهای کمی ۸ نفره هستند و بیش‌تر خانوارها ۳ نفره هستند. بعد از آن خانوارهای ۴ نفره

بیش‌ترین تکرار را دارند. اعداد سطر دوم در جدول صفحهٔ قبل نشان می‌دهد که هریک از اعداد سطر اول چند بار تکرار شده‌اند. این تکرارها را **فراوانی مطلق** می‌گوییم.

فراوانی مطلق داده x_i برابر تعداد دفعاتی است که آن داده تکرار شده است. فراوانی مطلق دادهٔ x_i را با نماد f_i نشان می‌دهیم.

در مثال صفحهٔ قبل $f_1 = ۴$ ، $f_۵ = ۶$ و $f_۸ = ۱$ است. در این جدول می‌بینیم که $f_1 + f_2 + ... + f_8 = ۴۰$. آیا می‌توانید توضیح دهید که این برابری همواره برقرار است؟ یعنی مجموع فراوانی‌ها برابر تعداد افراد و یا اشیایی است که مورد مطالعه قرار گرفته‌اند. البته می‌توانستیم بنابر تقسیمات رایج، خانوارها را به‌صورت زیر دسته‌بندی کنیم :

ــ خانوار کم‌جمعیت ؛ یعنی خانواری که جمعیت آن یک یا دونفر باشد،

ــ خانوار با جمعیت متوسط ؛ یعنی خانواری که جمعیت آن ۳ ، ۴ و یا ۵ نفر باشد،

ــ خانوار پرجمعیت ؛ یعنی خانواری که جمعیت آن بیش از ۵ نفر باشد.

در این صورت می‌توانستیم جدول زیر را برای داده‌های بالا تنظیم کنیم :

نوع خانوار	کم‌جمعیت	باجمعیت متوسط	پرجمعیت	جمع
تعداد خانوارها	۱۱	۲۳	۶	۴۰

این جدول نشان می‌دهد که بیش‌تر خانوارها دارای جمعیت متوسط هستند و تعداد خانوارهای پرجمعیت از تعداد خانوارهای کم‌جمعیت، کم‌تر است.

مجدداً در این جدول هم اعداد ۱۱ ، ۲۳ و ۶ را فراوانی مطلق می‌گوییم. در نمونه تحت بررسی خانوادهٔ کم‌جمعیت ۱۱ بار تکرار شده است.

برای توزیع کالابرگ جدول اول مناسب است و حال آن‌که در ساخت و ساز ساختمان‌های مسکونی اگر چه جدول اول هم می‌تواند مفید باشد ولی معمولاً جدول دوم مبنا قرار می‌گیرد.



دسته‌بندی

اولین اقدامی که در مطالعهٔ یک جامعه براساس داده‌ها انجام می‌دهیم، آن است که ببینیم آیا می‌توان جامعه را به چند دسته جدا از هم تفکیک کرد؟ البته ما هر مجموعه‌ای را می‌توانیم به چند دسته تقسیم کنیم، ولی منظور ما از تشکیل دسته آن است که اعضای موجود در یک دسته، آشنا و یک‌دست باشند. در واقع یک‌دسته همگن، همگون و متناسب داشته باشیم و تفاوت بسیار در آن دیده نشود. مثلاً از نظر آب و هوا، ایران را می‌توان به چند ناحیه مشخص تقسیم کرد که در هریک از این نواحی آب و هوا یکسان است. به‌طور مثال شمال ایران، استان‌های مازندران و گیلان علی‌رغم وسعت زیاد آن، دارای آب و هوای تقریباً یکسان است. از غرب گیلان تا شرق مازندران تفاوت زیادی در آب وهوا ملاحظه نخواهید کرد. شهرهای حاشیه کویر ایران نیز از لحاظ آب و هوا یکسانند. نواحی جنوبی ایران نیز آب و هوای خاص خود را دارد. غرب ایران آب و هوایی دارد که با آب و هوای شرق ایران متفاوت است. در هرحال ملاحظه می‌شود که ایران از لحاظ آب و هوا به چند دستهٔ مشخص تقسیم می‌شود. به‌همین ترتیب مردم ایران را می‌توان به نژادها و زبان‌های چندی دسته‌بندی کرد. این دسته‌بندی‌ها، جامعه ما را به چند جزء کوچک‌تر که مثل هم هستند تقسیم می‌کند. پس از این دسته‌بندی، مسئله مهم یافتن اندازهٔ این دسته‌هاست و مطالعهٔ جامعه روال آسان‌تری پیدا می‌کند. پس از این مقدمه نسبتاً طولانی به معرفی جدول فراوانی می‌پردازیم.

در مطالعهٔ متغیرهای گسسته جدول فراوانی بدون دسته‌بندی کارآمد است، ولی اگر متغیر پیوسته باشد این روش برای منظم کردن داده‌ها عملی نیست. لذا برای تنظیم داده‌ها در این قبیل موارد مراحل زیر را طی می‌کنیم :

دامنه تغییرات: اولین نکته‌ای که در این مرحله مدنظر قرار می‌گیرد آن است که ببینیم آیا متغیر در فاصله زیادی تغییر می‌کند یا در فاصله کمی قرار می‌گیرد. طول بازه‌ای را که متغیر در آن تغییر می‌کند «**دامنه تغییرات**» می‌گوییم. برای بیان روشن‌تر، فرض کنید :

کوچک‌ترین داده = a و بزرگ‌ترین داده = b

در این‌صورت تفاضل a از b یعنی $b-a$ را **دامنهٔ تغییرات** می‌گویند و با علامت R نشان می‌دهیم.

$$R=b-a$$

بزرگی دامنهٔ تغییرات نشان‌دهندهٔ تفاوت زیاد در جامعه است. هرچقدر این دامنه کمتر باشد افراد جامعه از لحاظ این متغیر به‌هم نزدیک‌ترند. اگر دامنهٔ تغییرات صفر باشد، تمام افراد باهم برابر و یکسانند.

در زیر سن چند تن از معلمان یک دبیرستان آمده است :

۳۰ ، ۴۵ ، ۶۲ ، ۳۷ ، ۴۲ ، ۴۸

دامنه تغییرات را در این داده‌ها محاسبه می‌کنیم :

$a=۳۰$ و $b=۶۲$

پس : $R=b-a=۶۲-۳۰=۳۲$

فعالیت

دامنه تغییرات قد را در بین همکلاسی‌های خود حساب کنید.

در فعالیت بالا ممکن است این دامنه تغییرات را براساس قد تمام همکلاسی‌های خود حساب کنید (منظور فعالیت نیز همین است) ولی ممکن است شما فکر کنید که ما یک نمونه ۸ تایی از این جامعه داریم اگر دامنه تغییرات را در نمونه حساب کنیم چه خواهد شد؟ چه ارتباطی بین این دو دامنهٔ تغییرات وجود دارد؟ برای روشن شدن مطلب به سؤال‌های زیر پاسخ دهید.

۱ــ از این دو دامنهٔ تغییرات کدام بزرگ‌تر و کدام کوچک‌تر است؟

۲ــ آیا زیاد بودن دامنه تغییرات در نمونه، زیاد بودن دامنهٔ تغییرات در جامعه را نتیجه می‌دهد؟

۳ــ آیا کوچک بودن دامنه تغییرات در نمونه، کوچک بودن دامنه تغییرات در جامعه را نتیجه می‌دهد؟

اگر دامنه تغییرات به‌نظر کوچک باشد، دیگر نیازی به دسته‌بندی جامعه نیست زیرا کوچک بودن دامنه تغییرات می‌گوید که اندازه‌های اعضا به هم نزدیک‌اند و اختلاف آن‌ها قابل ملاحظه نیست پس با یک جامعه کاملاً یک‌دست سروکار داریم. اما در مواردی که تغییرات زیاد است، جامعه را نمی‌توانیم یک‌دست فرض کنیم و لذا باید آن را دسته‌بندی کنیم.

بحث کلاسی

ــ در کلاس خود متغیری را مثال بزنید که کلاس از این‌نظر یک‌دست باشد.

ــ در کلاس خود متغیری را مثال بزنید که کلاس از نظر این متغیر یک‌دست نباشد.

حال فرض کنید دامنهٔ تغییرات زیاد باشد مانند میزان بارندگی در شهرهای مختلف ایران. این متغیر در کشور ما دامنهٔ تغییرات زیادی دارد. دامنهٔ تغییرات بارندگی در شمال کم است و به‌همین ترتیب در شهرهای حاشیه کویر. ابتدا دامنهٔ تغییرات را به چند قسمت کوچک‌تر تقسیم می‌کنیم.

سپس براساس این قسمت‌بندی، داده‌ها را در دسته‌های مختلف توزیع می‌کنیم. تعداد دسته‌ها باید به گونه‌ای باشد که در هر دسته تغییرات کم باشد.

برای روشن بودن مطالب و ادامه بحث، موضوع را با یک مثال پی می‌گیریم.



داده‌های زیر یک نمونهٔ ۲۵تایی از نمرات ریاضی یک کلاس است.

۱۷/۵، ۱۵/۵، ۱۱، ۱۲/۲۵، ۱۵/۵، ۱۱/۵، ۱۰، ۱۴/۲۵، ۱۳/۵، ۱۵، ۱۹، ۱۸/۵، ۲۰، ۱۴، ۱۳، ۱۲/۵، ۱۸، ۱۸/۵، ۱۷/۲۵، ۱۶/۵، ۱۱، ۱۱/۵، ۱۴/۵، ۱۳/۵، ۱۵/۵

در اینجا داریم :

$$a=۱۰ \text{ و } b=۲۰ \quad , \quad R=۲۰-۱۰=\boxed{۱۰}$$

حال این دامنه تغییرات را چهار قسمت می‌کنیم. انتخاب این تعداد قانون و قاعده خاصی ندارد بلکه توصیه‌هایی دربارهٔ آن هست که به‌تدریج به آن‌ها اشاره می‌شود. با این تصمیم ما داده را به چهار دسته [۱۷/۵,۲۰] و (۱۵,۱۷/۵] و (۱۲/۵,۱۵] و (۱۰,۱۲/۵] تقسیم کرده‌ایم. در اینجا (۱۲/۵,۱۵] یعنی اعدادی که از ۱۵ کوچکترند و بزرگتر یا مساوی ۱۲/۵ هستند. منظور ما از این تقسیم‌بندی آن است که مقادیر موجود در دستهٔ اول یعنی (۱۰,۱۲/۵] ، اگرچه **متفاوت‌اند ولی از نظر ما یکی هستند** و از تفاوت موجود بین آن‌ها می‌توان صرف‌نظر کرد. همین نکته یکی از توصیه‌های انتخاب تعداد دسته‌هاست. اگر از اختلاف بین ۱۰ و ۱۲/۵ نتوانیم چشم‌پوشی کنیم و این اختلاف به نظرمان بزرگ بیاید، باید فاصله‌ها را کوچک‌تر کنیم. پس باید تعداد آن‌ها را بیش‌تر کنیم. مثلاً داده‌ها را به جای چهار دسته به پنج دسته تقسیم کنیم.

مرکز دسته: داده‌ها را به‌نحوی دسته‌بندی کردیم که افرادی که در یک دسته قرار می‌گیرند از نظر اندازه متغیر مورد مطالعه تفاوت چندانی باهم نداشته باشند. در اصل ما می‌توانیم با کمی اغماض این افراد را دارای یک اندازه بدانیم. حال سؤال این است که این اندازه مشترک برای آن‌ها چه‌مقداری باشد، بهتر است. چون اختلاف‌ها را در هر دسته ناچیز فرض کرده‌ایم، هریک از مقادیر حاصل از آن دسته را می‌توانیم برای هریک از افراد آن دسته درنظر بگیریم ولی اگر بتوان کاری کرد که این یکسان‌سازی داده‌ها در دسته‌ها تأثیر کم‌تری در روند مطالعه بگذارند، مسلماً بهتر خواهد بود. از این‌رو مناسب‌ترین مقداری که می‌توانیم به هریک از افراد یک دسته نسبت دهیم، مقدار میانی و یا وسط دسته است. این مقدار را مرکز دسته می‌گوییم. اگر دستهٔ iام به‌صورت $[a_i, b_i)$ باشد مرکز آن را که با نماد x_i نشان می‌دهیم به‌صورت زیر تعریف می‌شود.

$$x_i = \frac{b_i + a_i}{۲}$$

تعریف : برای دسته $[a_i, b_i)$، a_i را **کران پایین**، b_i را **کران بالا** و x_i را **مرکز آن دسته** می‌نامیم.

در دستهٔ دوم $[۱۲/۵ , ۱۵)$ مرکز دسته را محاسبه می‌کنیم :

$$x_i = \frac{۱۲/۵+۱۵}{۲} = \frac{۲۷/۵}{۲} = ۱۳/۷۵$$

ما اگر داده‌ها را در مثلاً ۴ دسته، دسته‌بندی کرده باشیم، به جای تعداد زیادی از داده‌ها ما فقط با ۴ عدد که همان مرکز دسته‌ها باشند سروکار داریم. البته چون قرار ما بر این است که داده‌های داخل یک دسته را برداریم و به‌جای آن‌ها مرکز دسته را قرار دهیم، این اعداد تکراری خواهند بود.

داده‌های زیر نمرات آمار و مدلسازی یک نمونهٔ ۸ نفری از دانش‌آموزان یک کلاس هستند.

۱۶،۱۶/۵،۱۸/۵،۱۷،۱۷/۵،۱۸،۱۹/۵،۲۰

در این داده‌ها داریم :

$a = ۱۶$ و $b = ۲۰$ ، $R = ۲۰ - ۱۶ = ۴$

داده‌ها را در دو دسته، دسته‌بندی می‌کنیم. این دسته‌ها عبارتند از: $[۱۶,۱۸)$ ، $[۱۸,۲۰]$. در دسته اول، ۱۶ کران پایین دسته اول و ۱۸ کران بالای دسته اول می‌نامیم، به همین ترتیب در دستهٔ دوم ۱۸ کران پایین دستهٔ دوم و ۲۰ کران بالای دسته دوم می‌باشند. داده‌های موجود در هر دسته را در جدول زیر می‌نویسیم.

داده‌های هر دسته	مرکز دسته‌ها	دسته‌ها
۱۶،۱۶/۵،۱۷،۱۷/۵	۱۷	۱۶ ـ ۱۸
۱۸،۱۸/۵،۱۹/۵،۲۰	۱۹	۱۸ ـ ۲۰

حال از نقطه‌نظر ما داده‌هایی که در یک دسته قرار دارند، یکسان درنظر گرفته می‌شوند و مقدار مشترک آن‌ها همان مرکز دسته‌هاست. بنابراین، اطلاعات بالا به‌صورت زیر هموارتر می‌شود.

داده‌های هر دسته	مرکز دسته‌ها	دسته‌ها
۱۷/۱۷،۱۷/۱۷	۱۷	۱۶ ـ ۱۸
۱۹/۱۹ ،۱۹/۱۹	۱۹	۱۸ ـ ۲۰

پس در این مسئله ما با دو مقدار ۱۷ و ۱۹ سروکار داریم که هریک از آن‌ها به تعدادی تکرار شده‌اند. در این مثال ملاحظه می‌کنید که مرکز دسته ممکن است یکی از داده‌ها نباشد.

تفاضل دو کران پایین متوالی یا دو کران بالای متوالی را طول آن دسته می‌نامیم.



تمرین

۱ــ در مثال نمرات آمار و مدلسازی طول دسته را حساب کنید.

ــ همچنین تعداد تکرارهای مراکز دسته‌ها را حساب کنید.

ــ مجموع این تکرارها چقدر است؟ این مجموع برابر چه چیزی است؟

با توجه به تقسیمات آب و هوایی، ناحیه خشک به ناحیه‌ای گفته می‌شود که دارای بارش سالانه ۳۰ تا ۲۵۰ میلی‌متر است و ناحیه نیمه‌خشک ناحیه‌ای است که دارای بارش سالانه از ۲۵۰ تا ۶۰۰ میلی‌متر است و ناحیه مرطوب ناحیه‌ای است که بارش سالانه آن از ۶۰۰ تا ۲۰۰۰ میلی‌متر است. به سؤالهای زیر پاسخ دهید :

ــ چند درصد مراکز استان‌های ایران در ناحیه خشک و نیمه مرطوب قرار دارند؟

ــ حداکثر اختلاف میزان بارندگی در مراکز استان‌ها چقدر است؟

ـ جدول فراوانی را برای داده‌های زیر تشکیل دهید :

ـ اگر داده‌ها را در چهار دسته قرار دهیم، کران بالا و پایین دسته دوم را تعیین کنید.

ـ حداکثر اختلاف میزان بارندگی در هر دسته چقدر است؟ آیا این مقدار برای تمام دسته‌ها یکسان است؟

ـ در دسته سوم چند شهر قرار دارد؟ میزان بارندگی شهرهای این دسته حدوداً چقدر است؟ و کدام دسته بیش‌ترین شهرها را دارد؟

مقداربارندگی سالانه مراکز استان‌ها **(میلی‌متر)**

مرکز استان	سال
تهران	۳۲۸
اراک	۴۷۵
رشت	۱۷۳۴
ساری	ـ
تبریز	۳۲۶
ارومیه	۴۲۶
کرمانشاه	۴۰۴
اهواز	۲۰۱
شیراز	۵۸۲
کرمان	۳۰۹
مشهد	۳۶۹
اصفهان	۱۹۸
زاهدان	۹۹
سنندج	۳۶۹
همدان	۴۲۵
شهرکرد	۴۲۶
خرم‌آباد	۷۶۲
ایلام	۶۳۵
یاسوج	۱۲۶۵
بوشهر	۲۶۰
زنجان	۳۱۲
سمنان	۱۶۰
یزد	۱۰۵
بندرعباس	۳۹۱

سالنامهٔ مرکز آمار سال ۱۳۷۱

انواع فراوانی

فراوانی مطلق: داده‌های زیر را که مربوط به ضریب هوشی ۷ دانش‌آموز پسر است در نظر بگیرید :

۱۱۲، ۱۱۷، ۱۱۶، ۱۱۲، ۱۱۱، ۱۱۷، ۱۱۲

برخی از این داده‌ها تکرار شده‌اند. در جدول زیر این داده‌ها را با تعداد تکرارهای آن نوشته‌ایم :

تعداد تکرارها	داده‌ها
۱	۱۱۱
۳	۱۱۲
۱	۱۱۶
۲	۱۱۷

تکرار هر داده را **فراوانی مطلق** آن داده و جدول فوق را جدول فراوانی می‌گوییم. در مثال بالا فراوانی مطلق ۱۱۱ برابر ۱ و فراوانی مطلق ۱۱۲ برابر ۳ است. به همین ترتیب فراوانی‌های مطلق ۱۱۶ و ۱۱۷ به ترتیب برابر ۱ و ۲ می‌باشد.

در دسته‌بندی داده‌ها قرار گذاشتیم تمام داده‌های واقع در یک دسته را برابر مرکز آن دسته درنظر بگیریم. بنابراین مرکز دسته‌ها به تعداد اعضایی که در آن دسته قرار دارند تکرار خواهند شد. یعنی فراوانی مرکز دسته‌ها برابر تعداد اعضایی است که در آن دسته قرار گرفته‌اند. این فراوانی را فراوانی مطلق آن دسته می‌گوییم.



فراوانی مطلق دستهٔ iام را که با f_i نشان می‌دهیم برابر تعداد اعضایی است که در دسته iام قرار گرفته‌اند.

فعالیت

جدول فراوانی را برای وزن همکلاسی‌هایتان تشکیل دهید و مجموع فراوانی‌های مطلق را حساب کنید و تحقیق کنید این مجموع برابرچیست؟

فراوانی نسبی: دو جدول زیر را درنظر بگیرید :

دسته‌ها	مرکز دسته‌ها	فراوانی مطلق
۱_۳	۲	۶
۳_۵	۴	۱۰
۵_۷	۶	۴
		۲۰

دسته‌ها	مرکز دسته‌ها	فراوانی مطلق
۱_۳	۲	۱۹
۳_۵	۴	۲۵
۵_۷	۶	۱۶
		۶۰

اولین سؤالی که ممکن است به‌نظر خواننده بیاید این است که این دو جدول چگونه حاصل شده‌اند؟ معمولاً یکی از مطالعات آماری مقایسه دو نوع رفتار، بر روی یک مجموعه است. مثلاً فرض کنید دو روش تدریس داریم که روش اول را برای ۲۰ نفر و روش دوم را برای ۶۰ نفر دانش‌آموزان هم‌سطح اعمال کرده‌ایم، از این افراد امتحانی یکسان گرفته‌ایم که نمرات در این امتحان می‌تواند از ۱ تا ۷ تغییر کند. کسانی که نمره آن‌ها در [۱,۳) است ضعیف و کسانی که نمرهٔ آن‌ها در [۳,۵) است





متوسط و نمره کسانی که در [۵,۷] است قوی ارزیابی می‌شوند. حال می‌خواهیم عمل‌کرد این دو روش آموزش را با هم مقایسه کنیم. اگر تعداد دانش‌آموزان در دو کلاس برابر بودند، به‌سادگی می‌توانستیم این مقایسه را انجام دهیم و معلوم کنیم که با کدام روش آموزش موفق‌تر بوده است. ولی از آنجایی که تعداد دانش‌آموزان کلاس‌ها برابر نیستند، به‌سادگی نمی‌توانیم این مقایسه را به انجام برسانیم. برای حل این مشکل با فرض این که عمل‌کرد روش آموزش بستگی به تعداد دانش‌آموزان کلاس ندارد (چون اگر بدانیم که بستگی دارد این مطالعه نمی‌تواند مبنایی برای مقایسه موفقیت دو روش آموزش باشد. در این صورت توضیح دهید که تعداد افراد دو کلاس چگونه باید باشند؟) فرض می‌کنیم تعداد دانش‌آموزان هر کلاس برابر ۱۰۰ نفر باشد و با یک تناسب ساده تعداد دانش‌آموزان ضعیف، متوسط، قوی را در هر کلاس حساب می‌کنیم. مثلاً در دستهٔ اول از کلاس اول محاسبات زیر را داریم:

تعداد دانش‌آموزان ضعیف	تعداد در کلاس
۶	۲۰
x	۱۰۰

$$x = ۱۰۰ \times \frac{۶}{۲۰} = ۳۰$$

همان‌طوری که قبلاً دیده‌اید، ۳۰ درصد دانش‌آموزان کلاس اول ضعیف ارزیابی شده‌اند. به همین ترتیب می‌توانیم برای کلاس دیگر و دسته‌های بعدی روش بالا را تکرار کنیم و درنتیجه جدول‌های زیر حاصل می‌شود.

دسته‌ها	مرکز دسته‌ها	فراوانی مطلق	درصد
۱ـ۳	۲	۶	۳۰
۳ـ۵	۴	۱۰	۵۰
۵ـ۷	۶	۴	۲۰
جمع	ـ	۲۰	۱۰۰

دسته‌ها	مرکز دسته‌ها	فراوانی مطلق	درصد
۱ـ۳	۲	۱۹	۳۱/۶
۳ـ۵	۴	۲۵	۴۱/۷
۵ـ۷	۶	۱۶	۲۶/۷
جمع	ـ	۶۰	۱۰۰

چون منظور، مقایسهٔ دو آزمون بود جدول صفحهٔ بعد را که خلاصهٔ اطلاعات را دربر دارد درنظر می‌گیریم.

	درصد کلاس اول	درصد کلاس دوم
ضعیف	۳۰	۳۱/۶
متوسط	۵۰	۴۱/۷
قوی	۲۰	۲۶/۷

از مقایسهٔ ستون‌های این دو جدول ملاحظه می‌شود که ترکیب دانش‌آموزان ضعیف در هر دو روش تقریباً ثابت بوده است (در حدود ۳۰، ۳۱ درصد) ولی روش دوم روی دانش‌آموزان متوسط موفق‌تر عمل کرده است. زیرا از درصد متوسط‌های کلاس اول کاسته شده ولی به درصد دانش‌آموزان قوی کلاس دوم اضافه شده است.

دیدیم که برای مقایسه لازم است تأثیر اختلاف نمونه را از بین ببریم برای این منظور تصمیم گرفتیم که اندازهٔ نمونه را ۱۰۰ فرض کنیم که در این‌صورت درصد فراوانی‌ها را حساب کردیم و برای محاسبهٔ این درصد دیدیم که باید عبارت $\frac{\text{فراوانی}}{\text{تعداد نمونه}}$ را در ۱۰۰ ضرب کنیم که آن را درصد فراوانی نسبی و نسبت $\frac{\text{فراوانی}}{\text{تعداد نمونه}}$ را **فراوانی نسبی** می‌گوییم.

اگر f_i فراوانی دسته iام و تعداد داده‌ها n باشد کسر $\frac{f_i}{n}$ را فراوانی نسبی دسته iام می‌گوییم.

جدول فراوانی موضوع مورد بحث به‌صورت زیر تکمیل می‌شود.

دسته‌ها	مرکز دسته‌ها	فراوانی مطلق	درصد فراوانی نسبی	فراوانی نسبی
۱ـ۳	۲	۶	۳۰	۰/۳
۳ـ۵	۴	۱۰	۵۰	۰/۵
۵ـ۷	۶	۴	۲۰	۰/۲
جمع		۲۰	۱۰۰	۱

فعالیت

با همکاری دانش‌آموزان جدول فراوانی نمرات تمام دانش‌آموزان را برای هر درس تهیه کنید و در کلاس نصب نمایید.

فراوانی تجمعی: جدول فراوانی روبرو تعداد کالاهایی را که پس از فروش برای تعمیرات اولیه به کارخانهٔ سازنده عودت داده شده‌اند، نشان می‌دهد. در این جدول ۳ کالا در فاصلهٔ ۱۰ تا ۱۳ ماه و ۶ کالا در فاصلهٔ ۱۳ تا ۱۶ ماه و ۷ کالا در فاصله ۱۶ تا ۱۹ ماه و ۴ کالا در فاصلهٔ ۱۹ تا ۲۲ ماه عودت داده شده‌اند.

فراوانی	دسته‌ها
۳	۱۰ـ۱۳
۶	۱۳ـ۱۶
۷	۱۶ـ۱۹
۴	۱۹ـ۲۲

چند کالا در فاصلهٔ ۱۰ تا ۱۹ ماه عودت داده شده است؟ فاصله ۱۰ تا ۱۹ به دسته‌های ۱۰ تا ۱۳، ۱۳ تا ۱۶ و ۱۶ تا ۱۹ تقسیم شده‌است. بنابراین تعداد کل کالاهایی که در فاصلهٔ ۱۰ تا ۱۹ عودت داده شده‌اند عبارت خواهد بود از :

۳+۶+۷=۱۶

این عدد را **فراوانی تجمعی** دسته ۱۶ـ۱۹ می‌گویند.

فراوانی تجمعی هر دسته برابر تعداد اشیایی است که مقدار آن‌ها از کران بالای آن دسته کمتراند.

مثلاً فراوانی تجمعی دستهٔ اول عبارت است از تعدادی که از ۱۳ کم‌ترند و این برابر همان فراوانی مطلق دستهٔ اول است. فراوانی تجمعی دسته دوم یعنی تعدادی که از ۱۶ کم‌ترند و این برابر ۶+۳ است.

به این ترتیب می‌توانیم فراوانی تجمعی دسته‌ها را حساب کنیم. معمولاً ستونی به جدول فراوانی تحت عنوان فراوانی تجمعی اضافه می‌کنند. در مثال بالا جدول فراوانی را می‌توان به‌صورت روبرو تکمیل کرد.

فراوانی تجمعی	فراوانی‌ها	دسته‌ها
۳	۳	۱۰ـ۱۳
۹	۶	۱۳ـ۱۶
۱۶	۷	۱۶ـ۱۹
[۲۰]	۴	۱۹ـ۲۲
	[۲۰]	جمع

سؤال: در این جدول آیا همیشه دو عددی را که در ☐ نشان داده‌ایم باهم برابرند؟ چه دلیلی برای پاسخ خود دارید؟

محاسبهٔ فراوانی تجمعی هر دسته، به‌صورت زیر است :

فراوانی تجمعی هر دسته برابر مجموع فراوانی‌های دسته‌های قبل و فراوانی آن دسته است.

چند تذکر دربارهٔ تشکیل جدول فراوانی

گفتیم که با بررسی اولیه در داده‌ها، با توجه به اندازهٔ نمونه و بزرگی یا کوچکی دامنهٔ تغییرات، تصمیم می‌گیریم که داده‌ها را در تعدادی دسته، دسته‌بندی کنیم، اگر این تعداد برابر k باشد با تقسیم R، دامنه تغییرات بر k، طول دسته‌ها به‌دست می‌آید. در بعضی مواقع ممکن است این دو عدد خارج قسمت خوبی به ما ندهند (اعشار زیاد داشته باشد و یا اصلاً قابل تقسیم نباشند). مثلاً فرض کنید $R = ۲۹$ و $k = ۷$ در این صورت طول دسته‌ها عبارت است از :

$$C = \frac{R}{k} = \frac{۲۹}{۷} = ۴/۱۴۲۸$$

ابتدا توجه می‌کنیم که وقتی به اعشار متوسل می‌شویم، منظور ما این است که می‌خواهیم اختلاف را در حد یک‌دهم، یک‌صدم و یا حتی یک‌هزارم کنترل کنیم. اگر C را به‌صورت ۴/۱۴ درنظر بگیریم حداکثر اختلاف ۴/۱۴ با $\frac{۲۹}{۷}$ برابر یک‌صدم است و اگر به‌صورت ۴/۱۴۲ در نظر بگیریم حداکثر اختلاف برابر یک‌هزارم است، آیا واقعاً ما به‌این همه دقت نیاز داریم؟ توجه کنید که ایدهٔ اصلی در تشکیل جدول فراوانی ازبین بردن اختلافات جزئی و یک کاسه کردن داده‌های نزدیک به هم است.

حال اگر بخواهیم در تعیین طول دسته‌ها به اعشار زیاد از حد متوسل شویم این ایده اصلی تشکیل جدول فراوانی را نادیده گرفته‌ایم، از این‌رو با درنظر گرفتن یک مقدار تقریبی مناسب برای C مسئله اعشار زیاد را حل می‌کنیم. ممکن است پیشنهاد کنید که در مثال بالا بهترین مقداری که برای C درنظر می‌گیریم عدد ۴ باشد، توجه کنید ۷ دسته به‌طول ۴، فاصله‌ای به طول ۲۸ را خواهند پوشاند که این از دامنهٔ تغییرات کم‌تر است. **دسته‌ها باید به‌نوعی باشند که مجموع طول آن‌ها از دامنه تغییرات کم‌تر نباشد.**

در این مثال شما می‌توانید طول دسته‌ها را هریک از مقادیر ۴/۲ یا ۴/۵ درنظر بگیرید، در این صورت فاصله‌ای که با $C = ۴/۲$ پوشانده می‌شود برابر $۷ \times ۴/۲ = ۲۹/۴$ که به اندازهٔ ۰/۴ از R بیش‌تر است. حال اگر دستهٔ اول را شما از کوچک‌ترین داده شروع کنید دستهٔ آخر به عددی ختم خواهد شد که ۰/۴ از بزرگ‌ترین داده بیش‌تر است که البته اشکال چندانی ندارد ولی در ریاضیات توجه به زیبایی مهم است.

از زیباترین مفاهیم ریاضی موضوع تقارن است. از این‌رو برای زیبایی بیشتر بهتر است این مقدار را به دو قسمت مساوی تقسیم کنیم و اولین دسته را از ۰/۲ کم‌تر شروع کرده تا آخرین دسته به ۰/۲ بیش‌تر ختم شود.

در یک امتحان تستی که شامل ۳۰ سؤال بوده است یک نمونهٔ ۲۰ نفری از دانش‌آموزان نمرات زیر را کسب کرده‌اند :

۱۲، ۱۹، ۱۴، ۲۷، ۱۷، ۲۵، ۵، ۱۲، ۳۰، ۲۹، ۲۸، ۲۹، ۲۵، ۱۳، ۲۴، ۷، ۱۶، ۹، ۸، ۱

این داده‌ها را در هفت دسته، دسته‌بندی کنید.

در این داده‌ها داریم، $a=۱$ ، $b=۳۰$ پس $R=۲۹$ پس $C=\frac{۲۹}{۷}\approx ۴/۲$

در این صورت دسته‌ها عبارتند از :

فراوانی مطلق	مرکز دسته‌ها	دسته‌ها
۱	۲/۹	[۰/۸ ، ۵)
۴	۷/۱	[۵ ، ۹/۲)
۳	۱۱/۳	[۹/۲ ، ۱۳/۴)
۳	۱۵/۵	[۱۳/۴ ، ۱۷/۶)
۱	۱۹/۷	[۱۷/۶ ، ۲۱/۸)
۳	۲۳/۹	[۲۱/۸ ، ۲۶)
۵	۲۸/۱	[۲۶ ، ۳۰/۲]
۲۰	جمع	

اگر طول دسته‌ها را ۴/۵ بگیریم خواهیم داشت :

$$۴/۵\times ۷=۳۱/۵$$

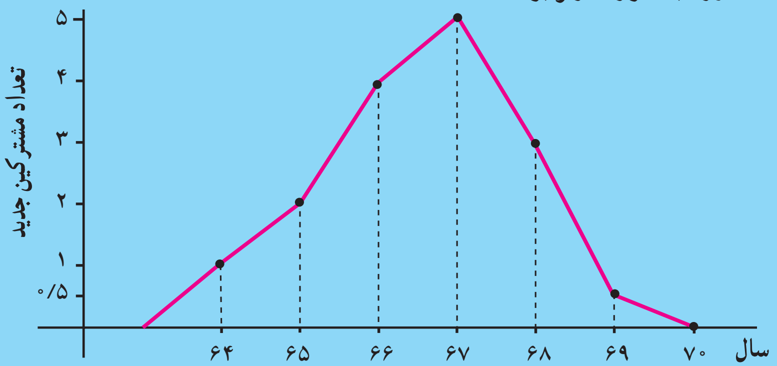
پس به اندازه $۳۱/۵-۲۹=۲/۵$ واحد از دامنه تغییرات زیادتر خواهیم داشت پس دستهٔ اول را به‌جای ۱ از $۱-\frac{۲/۵}{۲}=-۰/۲۵$ شروع می‌کنیم.

ملاحظه می‌کنید که این دو جدول تفاوت‌هایی در ستون فراوانی مطلق و مرکز دسته‌ها دارند ولی این اختلاف‌ها در تفسیر داده‌ها و نتیجه‌گیری‌های آماری تأثیری نخواهد داشت.

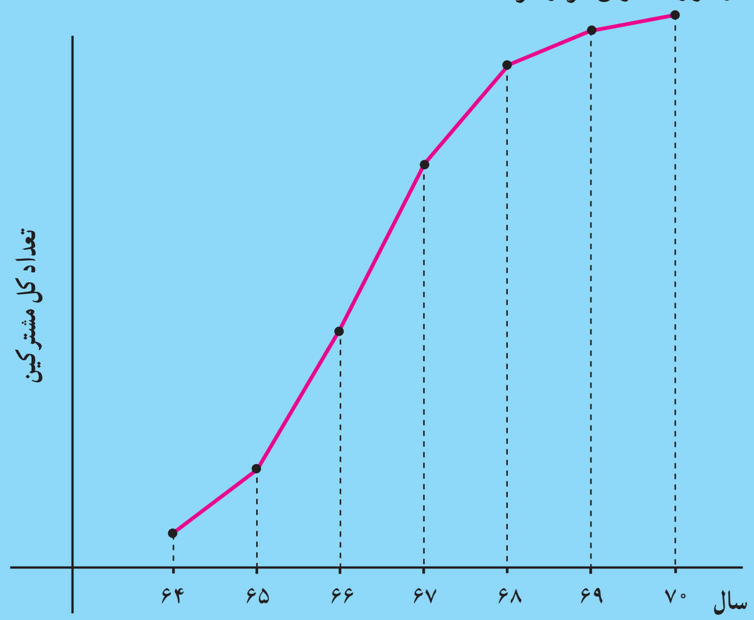
فراوانی	مرکز دسته‌ها	دسته‌ها	
۱	۲	۴/۲۵	-۰/۲۵
۳	۶/۵	۸/۷۵	۴/۲۵
۴	۱۱	۱۳/۲۵	۸/۷۵
۲	۱۵/۵	۱۷/۷۵	۱۳/۲۵
۲	۲۰	۲۲/۲۵	۱۷/۷۵
۳	۲۴/۵	۲۶/۷۵	۲۲/۲۵
۵	۲۹	۳۱/۲۵	۲۶/۷۵
۲۰	جمع		

بحث کلاسی

مدیر بانکی از هیأت مدیره برای بازدید شعبه دعوت کرد. این مدیر، تصمیم گرفت نموداری از وضعیت بانک به نمایش بگذارد. نموداری که او دربارهٔ تعداد مشترکین به‌دست آورد به‌صورت زیر بود.



این مدیر متوجه شد که از سال ۶۷ تعداد مشترکین جدید روبه کاهش گذاشته است و اتفاقاً این سال مصادف با سالی است که او مسئولیت این شعبه را پذیرفته بود. این وضعیت، وضعیت مناسبی نبود. او به‌فکر چاره افتاد و سرانجام نمودار زیر را رسم کرد و مورد تشویق قرار گرفت.



ــ می‌توانید بگویید برای رسم این نمودار از چه فراوانی استفاده شده است؟

ــ می‌توانید بگویید از کجا فراوانی‌های مطلق روبه کاهش گذاشته است؟

تمرین

۱ــ در دسته‌های جدول فراوانی و مرکز دسته‌ها داریم :

فاصله بین دو مرکز دسته متوالی = طول دسته

برابری بالا را با یک مثال نشان دهید. سپس با استفاده از این برابری جدول روبرو را تکمیل کنید.

۷ –	۱۱	
۱۱		
		۲۵

۲ــ می‌دانیم شکستن رمزها و کدها اهمیت زیادی در امنیت ملی دارد.

ساده‌ترین نوع رمزی کردن یک پیغام این است که هرحرف را با یک علامت جایگزین کنیم. معمولاً اولین قدم در کشف این نوع رمز بررسی فراوانی هر علامت در متن رمزنگاری شده است. زیرا معمولاً درصد حروفی که در متون یک زبان خاص به‌کار رفته‌اند مشخصه‌ای از آن زبان هستند. در جدول روبرو درصد فراوانی حروف در متون زبان انگلیسی نمایش داده شده‌اند.

حروف	درصد تکرار
E	۱۳
T	۹
A,O	۸
N	۷
I,R	۶/۵
S,H	۶
D	۴
L	۳/۵
C,U,M	۳
F,P,Y	۲
W,G,B	۱/۵
V	۱
K,X,J	۰/۵
Q,Z	۰/۲

جدول درصد فراوانی حروف در متون زبان فارسی را به‌دست بیاورید.

۳ــ می‌خواهید چاپگر کامپیوتر خریداری کنید. چاپگرهای سوزنی هنوز هم طرفداران بسیاری دارد. سرعت چاپ همیشه یک فاکتور مهم برای انتخاب بوده است. در مجلهٔ PC در مقاله ششمش بررسی سالانهٔ چاپگر (جلد هشتم شماره ۱۹) سرعت ۳۱۲ نوع مختلف چاپگر آمده است. یک نمونه تصادفی ۴۸ تایی از چاپگرهای سوزنی انتخاب شد. جدول صفحهٔ بعد سرعت این چاپگرها را برحسب کاراکتر در ثانیه نشان می‌دهد :

۲۴۷، ۱۹۵، ۱۴۵، ۹۱، ۱۰۰، ۱۱۱، ۲۵۸، ۹۹، ۵۳، ۶۰، ۱۲۴، ۱۳۳، ۱۷۳، ۴۶، ۱۱۵، ۹۳
۱۱۱، ۱۳۸، ۱۶۳، ۱۰۵، ۱۴۱، ۱۳۸، ۱۶۳، ۱۰۵، ۱۴۱، ۱۳۸، ۱۵۴، ۱۲۸، ۱۰۲، ۳۷، ۸۶، ۷۲
۱۷۳، ۱۵۷، ۱۵۰، ۱۳۳، ۸۹، ۸۹، ۱۴۷، ۱۷۳، ۱۱۱، ۱۵۷، ۱۴، ۱۵۰، ۱۰۰، ۱۳۱، ۱۳۶، ۱۵۴

الف ــ طول دسته‌ها را با در نظر گرفتن ۵ دسته به‌دست آورید.

ب ــ جدول فراوانی را با داشتن ۵ دسته تشکیل دهید. مرکز دسته‌ها را مشخص کنید.

۴ــ پرستاران یکی از بیمارستان‌های تهران اعتقاد دارند که نیاز به همکاران بیش‌تری در نوبت شب دارند. برای آن که حجم کارشان تخمین زده شود، نمونه‌ای از ۳۵ شب به‌طور تصادفی انتخاب شد و تعداد درخواست‌های بیماران در طول شب به‌قرار زیر ثبت شد :

۶۲، ۵۸، ۴۶، ۸۷، ۱۰۲، ۳۷، ۷۱، ۸۲، ۷۳، ۷۰، ۸۰، ۹۵، ۱۰۱، ۱۰۰، ۹۰، ۱۸، ۸۶، ۷۰، ۶۸
۸۸، ۷۱، ۷۵، ۸۶، ۸۳، ۷۴، ۶۰، ۷۷، ۹۰، ۸۶، ۸۴، ۸۵، ۶۹، ۷۰، ۹۲، ۶۳

جدول فراوانی داده‌های فوق را برای پنج دسته تشکیل دهید. اگر ۷۵ درخواست در شب کار سنگین درنظر گرفته شود، آیا فکر می‌کنید، درخواست این پرستاران به حق است؟

پروژه

در این پروژه می‌خواهیم بدانیم نوجوانان روزانه با چه گروه‌های سنی بیش‌تر سر و کار دارند.

موضوع: بررسی سن جامعهٔ خارج از مدرسه که شما در طی روز با آن‌ها سر و کار دارید (اعضای خانواده، فامیل، همسایگان، مغازه‌داران، راننده سرویس و ...) برای این بررسی حدود ۲۰ نفر از جامعهٔ بالا را در نظر بگیرید. جدول فراوانی سن آن‌ها را تشکیل دهید. به سؤالات زیر پاسخ دهید :

ــ داده‌های خود را چگونه جمع‌آوری نموده‌اید؟

ــ بیش‌ترین تعداد افرادی که با آن‌ها سر و کار داشته‌اید، درچه گروه سنی قرار دارند؟

ــ این مجموعه چه درصدی از افرادی را تشکیل می‌دهند که با آن‌ها سر و کار داشته‌اید؟

ــ توزیع سن را از طریق تعیین درصد فراوانی هریک از گروه‌های سنی مشخص کنید.

معرفی و آشنایی با نرم‌افزار آماری MINITAB



در سال‌های اخیر استفاده از کامپیوتر تقریباً در تمام جنبه‌های زندگی افزایش یافته است. استفاده از کامپیوتر زمان لازم برای تجزیه و تحلیل‌های کمّی را کاهش داده است.

در سال‌های اخیر نرم‌افزارهای آماری بسیاری به بازار وارد شده است. چهار نرم‌افزار اصلی آماری عبارتند از : BMDP ، MINITAB، SAS ، SPSS. در این بخش خلاصه‌ای از چگونگی کار کردن با MINITAB برای حل مسائل آماری آورده شده است. این نرم‌افزار تحت ویندوز ۹۵ عمل می‌کند.

MINITAB برای ویندوز

ابتدا وارد محیط ویندوز شوید : روی نمایهٔ آبی‌رنگ MINITAB دوبار کلیک کنید، تا وارد محیط MINITAB شوید. همین‌که وارد محیط MINITAB شدید، صفحهٔ کار MINITAB_UNTITLED را مشاهده می‌کنید، زیر آن لیست انتخاب PULL _ DOWN MENUS را مشاهده می‌کنید :

File, Edit, Manip, Calc, Stat, Graph, Editor, Window, Help

با کلیک کردن روی هریک از این انتخابات، لیست انتخابات در هریک نمایانگر می‌شود. به‌طور مثال اگر روی File کلیک کنید ؛ لیست زیر نمایان شده و شما می‌توانید هرکدام را انتخاب کنید :

File

NEW	Worksheet ...	Ctrl + N
Open	Worksheet ...	Ctrl + O
Merg	Worksheet ...	
Query Data base [ODBC]	...	
Save Worksheet		Ctrl + S
Save Worksheet As		
Worksheet Description		
Open Graph		
Save Window As		
Other Files		→
Print Window		Ctrl + P
Print Setup		
Restart Minitab		
Exit		

ورود داده‌ها

صفحه نمایش زیرِ لیست انتخابات به دو قسمت تقسیم شده است : Session و پنجرهٔ داده‌ها (Data Windows). از پنجره Session برای تایپ کردن فرمان‌های Minitab استفاده می‌شود. همچنین خروجی Minitab Output در این قسمت ظاهر می‌شود.

معمولاً داده‌ها را در قسمت پنجرهٔ داده‌ها (Data Windows) وارد می‌کنید. این قسمت شامل ستون‌هایی است که با C۱ و C۲ و ... نامگذاری شده‌اند. در Minitab داده‌ها همواره به شکل ستونی وارد می‌شوند. به‌طور مثال اگر داده‌های شما راجع به قد افراد است، این داده‌ها به‌صورت ستونی در یکی از ستون‌های C۱ یا C۲ یا... وارد می‌شوند. داده‌های کیفی را نیز می‌توان با وارد کردن یک نام یا مقدار در هر خانهٔ ستون ثبت نمود.

می‌توانید نام متغیر چون وزن، قد و ... را نیز وارد کنید. توجه کنید که هر خانهٔ ستون شماره‌گذاری شده است. هر دفعه که داده‌ای را وارد می‌کنید، کلید ENTER یا RETURN را فشار دهید یا کلید فلش به پایین ↓ را فشار دهید تا به خانه بعدی وارد شوید. (می‌توانید از کلیدهای جهت → ← ↓ ↑ استفاده نمایید) همچنین می‌توانید با استفاده از ماوس و حرکت نشان‌گر خانهٔ موردنظر را انتخاب کنید. زمانی که ورود تمام داده‌ها تمام شد می‌توانید از لیست انتخاب، تحلیل

مورد نیاز خود را انتخاب کنید.

تذکر: زمانی که داده‌ها را وارد می‌کنید از «،» استفاده نکنید برای مثال برای وارد کردن ۴۵،۷۶۳ باید ۴۵۷۶۳ تایپ شود.

مثال ۱: فرض کنید می‌خواهید داده‌های روبرو را در Minitab وارد کنید :

نام	نمرهٔ امتحان
علی	۱۲
حسن	۱۷
بهزاد	۱۴
امیر	۱۹
جلال	۲۰
حمید	۲۰
سعید	۱۹
احمد	۱۸

ــ ابتدا روی خانه C۱ کلیک کرده تایپ کنید NAME «نام» سپس به خانه زیر حرکت کنید (با استفاده از ماوس، کلیدهای جهتی، ENTER) و شروع به تایپ اسامی کنید. سپس روی C۲ کلیک کرده تایپ کنید SCORE «نمره» و ادامه دهید.

پنجرهٔ داده‌ها در پایان به شکل زیر درخواهد آمد :

Data

	C1–T	C2	C3	C4	C5	C6	C7
↓	Name	Score					
1	ALI	12					
2	HASSAN	17					
3	BEHZAD	14					
4	AMIR	19					
5	JALAL	20					
6	HAMEED	20					
7	SAEED	19					
8	AHMAD	18					
9							

اگر در وارد کردن داده‌ها اشتباهی صورت گرفت، نشانگر را روی خانهٔ موردنظر کلیک کرده کلید ENTER را فشار داده و سپس داده را وارد نمایید. اطلاعات جدید جایگزین اطلاعات قبلی می‌شود.

ذخیره کردن یک فایل Minitab

پس از ورود تمام داده‌ها می‌توانید فایل داده‌ها را برای استفاده در آینده ذخیره نمایید. به‌طور مثال می‌خواهید داده‌های مثال ۱ را ذخیره کنید :

قدم اول: روی لیست انتخاب، روی File کلیک کنید.

قدم دوم: روی Save Worksheet As کلیک کنید.

قدم سوم: نام فایل خود را در کادر File Name تایپ کنید. این نام معمولاً پسوند «MTW.» دارد به‌طور مثال فایل مثال را می‌توان SCORE. MTW نامید. سپس درایو مناسب را در قسمت Drives انتخاب کنید. برای مثال می‌توانید این فایل را در درایو A یا C و ... ذخیره نمایید، اگر بیش از یک دایرکتوری دارید، دایرکتوری موردنظر خود را انتخاب کرده سپس روی OK کلیک کنید. مطمئن شوید در کادر Save File As Type، لغت Minitab نشان داده شده باشد.

قدم چهارم: پس از انجام اعمال فوق روی OK کلیک کنید. در مثال بالا فایل فوق تحت نام SCORE.MTW ذخیره شد.

بازآوردن فایلی که قبلاً ذخیره شده است:

قدم اول: روی لیست انتخاب روی File کلیک کنید.

قدم دوم: روی Open Worksheet کلیک کنید.

قدم سوم: کادر Open Worksheet ظاهر می‌شود. درایوی را که فایل موردنظرتان در آن ذخیره شده است انتخاب کنید به‌طور مثال اگر درایو A انتخاب شود شما این عبارت را می‌بینید \:a و نام تمام فایل‌های Minitab در درایو A زیر کادر File Name نمایان می‌شوند. فایل موردنظر را انتخاب کرده و روی OK کلیک کنید.

قدم چهارم: پس از کلیک کردن روی OK، فایل موردنظر بازمی‌گردد.

ذخیره‌سازی فایل خروجی Minitab

فایل خروجی Minitab شامل پاسخ‌ها و نتایج تجزیه و تحلیل‌های انجام شده است. شما می‌توانید فایل خروجی را برای استفاده در آینده به‌عنوان فایل Text ذخیره کنید. فرض کنید در مثال ۱، یک سری محاسبات آماری انجام شده است و نتایج را می‌خواهید در فایلی به نام Score. Txt ذخیره کنید:

قدم اول: روی File کلیک کنید.

قدم دوم: روی Save Window As کلیک کنید.

قدم سوم: در کادر File Name ، نام فایل مورد نظر را تایپ کنید.

نام فایل خروجی معمولاً پسوند«TXT.» دارد که نشان‌دهنده فایل خروجی در Minitab است. درایو مورد نظر خود را در قسمت Drive انتخاب کرده و سپس در قسمت Directories، دایرکتوری موردنظرتان را انتخاب کنید.

اگر در فایل خروجی نمودار رسم شده است، مراحل فوق را برای ذخیره‌سازی انجام داده فقط به‌جای TXT. از «MGF.» استفاده کنید.

چاپ یک فایل خروجی Minitab

برای چاپ یک فایل خروجی شامل نمودار، قدم‌های زیر را انجام دهید :

قدم اول: روی File کلیک کنید.

قدم دوم: روی Print Window کلیک کنید.

قدم سوم: اگر تمام فایل خروجی را می‌خواهید چاپ کنید روی All کلیک کنید، در غیر این‌صورت شماره صفحات موردنظر برای چاپ را در قسمت FROM (از) و TO (تا) تایپ کنید.

قدم چهارم: روی OK کلیک کنید.

انتخاب نمونه با استفاده از Minitab

از این نرم‌افزار می‌توان در انتخاب نمونه‌ای از یک جامعه استفاده نمود. فرض کنید در مثال ۱، ازبین ۸ نفر می‌خواهید یک نمونهٔ ۳ نفری انتخاب کنید و این نمونه را در ستون‌های C۳ و C۴ ثبت کنید. توجه کنید که داده‌های جامعه در ستون‌های C۱ و C۲ ثبت شده‌اند :

قدم اول: روی Cal در لیست انتخاب کلیک کنید.

قدم دوم: روی Random Data کلیک کنید.

قدم سوم: روی Sample From Column کلیک کنید.

قدم چهارم: در کادرِ کنار Sample ، تایپ کنید : ۳ و در کادر Samples تایپ کنید :C۱ ــ C۲ . این دستور به این معنی است که سه ردیف از داده‌های وارد شده در ستون‌هایC۱ و C۲ را انتخاب کنید. سپس در کادر Store Sample in تایپ کنید C۴ ــ C۳ . اگر تکرار در نمونه‌تان مجاز است کنار کادر Sample With Replacement علامت بزنید.

قدم پنجم: روی OK کلیک کنید.

خروج از Minitab

قبل از آن که از Minitab خارج شوید، مطمئن شوید تا تمام فایل‌های داده‌ها و خروجی ذخیره شده‌اند و سپس چنین عمل کنید :

قدم اول: روی File کلیک کنید.

قدم دوم: روی Exit کلیک کنید.

قدم سوم: به سؤالات Minitab پاسخ دهید.

فرمان‌های زبانی Minitab

به جای استفاده از ویندوز می‌توان از فرمان‌های زبانی Minitab استفاده نمود. پس از ورود به محیط Minitab نشان‌گر روی Data Window خواهد بود که نشان‌دهنده این امر است که شما می‌توانید از Windows استفاده کنید. برای تغییر این وضعیت چنین عمل کنید :

قدم اول: در کادر Session Window کلیک کنید. نشان‌گر از Data Window به Session Window انتقال پیدا می‌کند.

قدم دوم: روی Editor در لیست انتخاب کلیک کنید.

قدم سوم: روی Enable Command Language کلیک کنید.

پس از انجام این عملیات پیغام : MTB> نمایان می‌شود. که Minitab Prompt خوانده می‌شود. در این حالت کامپیوتر آماده دریافت فرمان‌های Minitab می‌باشد. پس از تایپ فرمان‌های Minitab پس از Prompt کلید Enter یا Return را فشار دهید. برای بازگشتن به وضعیت Data Window ، روی لیست انتخاب Editor را انتخاب کرده روی Disable Command Language کلیک کنید.

ورود داده‌ها

مثال ۲: فرض کنید می‌خواهید داده‌های مثال ۱ را با استفاده از فرمان‌های Minitab وارد کنید. می‌توانید از فرمان‌های SET یا READ استفاده کنید. از SET برای ورود داده‌ها یکی‌یکی استفاده می‌شود. از READ برای ورود داده‌های یک متغیر یا متغیرهای بیش‌تری استفاده می‌شود.

فرمان Minitab که با Note شروع می‌شود فقط برای اطلاع رساندن به کاربر است و روی آن عمل نمی‌شود.

```
MTB > NOTE : DATA ON NAMES AND
             SCORES OF 8 STUDENTS
```

ورود داده‌ها در Minitab با استفاده از فرمان‌های زبانی Minitab:

این فرمان ستون C۱ را ایجاد می‌کند وجود; در انتهای فرمان اشاره بر این دارد که فرمان ادامه دارد و اطلاعات بیش‌تری وارد خواهد شد. به همین جهت PROMPT ملاحظه شده < SUBC خواهد بود و یعنی زیر فرمان. FORMAT (A6) نشان‌دهنده این است که یک متغیر اسمی حداکثر با ۶ کاراکتر داریم. توجه داشته باشید طولانی‌ترین نام ۶ کاراکتر دارد.

```
MTB > SET C1 ;
SUBC > FORMAT (A6).
```

نقطه در پایان فرمان نشان‌دهنده این است که اطلاعات پایان یافته.

DATA > ALI ↵

DATA > HASSAN ↵

DATA > BEHZAD ↵

DATA > AMIR ↵

DATA > JALAL ↵

DATA > HAMEED ↵

DATA > SAEED ↵

DATA > AHMAD ↵

DATA > END

MTB > SET C2

DATA > 12 17 14 19 20 20 19 18

DATA > END

پایان ستون C۲

پایان ستون C۱

ستون C۲ را آماده می‌کند. در این‌جا نیازی به «;» نمی‌باشد چون داده‌ها عدد هستند اعداد را می‌توان در یک ردیف نوشت ولی نام‌ها باید یکی‌یکی وارد شوند.

شما می‌توانید ستون‌های فوق را با استفاده از فرمان NAME، نام‌گذاری کنید :

MTB > NAME C1 'NAME'

MTB > NAME C2 'SCORE'

فرمان خروجی به قرار زیر است :

MTB > PRINT C1-C2

MTB > PRINT 'NAME' 'SCORE'

ذخیره‌سازی فایل ورودی

اگر تصمیم دارید از داده‌های وارد شده در Minitab بعدها استفاده کنید، باید آن‌ها را قبل از پایان دادن کار با Minitab ، ذخیره کنید. برای ذخیره کردن فایل ورودی ابتدا باید فایل را نام‌گذاری کنید. فرض کنید می‌خواهید نام «SCORE» را به فایل موجود در مثال ۲ نسبت دهید. از هریک از دو فرمان زیر می‌توان برای انجام این کار استفاده نمود. از این فرامین بسته به این که می‌خواهید روی درایو C یا A ذخیره کنید، استفاده کنید. توجه داشته باشید که نام فایل باید دربین ' ' قرار گیرد.

این فرمان فایل

MTB > SAVE 'SCORE'

SCORE. MTW را روی دیسک سخت ذخیره می‌کند.

این فرمان فایل

MTB > SAVE ‘ A : SCORE’

SCORE. MTW را روی دیسک Floppy در درایو A ذخیره می‌کند.

اگر خودتان پسوندی برای SCORE انتخاب نکرده باشید، Minitab پسوند MTW. را به نام فایل اختصاص می‌دهد.

مثال ۳: فایل SCORE را در دایرکتوری به نام STAT در درایو C ذخیره کنید :

MTB > SAVE ‘C: \ STAT \ SCOR’

بازآوری فایل‌های ورودی ذخیره شده در Minitab

با استفاده از فرمان RETRIEVE فایل ورودی را که قبلاً ذخیره شده است می‌توانید فعال کنید.

از این فرمان زمانی استفاده می‌شود که فایل روی درایو C در دیسک سخت ذخیره شده باشد.

MTB > RETRIEVE ‘SCORE’

از این فرمان زمانی استفاده می‌شود که فایل روی درایو A روی Floppy ذخیره شده باشد.

MTB > RETRIEVE ‘A: SCORE’

اگر فایل روی دایرکتوری خاصی در درایو C ذخیره شده باشد، باید دایرکتوری را مشخص نموده به‌طور مثال برای بازآوری فایل ورودی مثال ۳ چنین عمل کنید :

MTB > RETRIEVE ‘C :\ STAT \ SCORE’

در MINITAB ، فرمان با استفاده از چهار حرف اول شناخته می‌شود پس به‌جای استفاده از RETRIEVE می‌توان از RETR استفاده نمود. فراموش نکنید نام فایل باید همواره بین ’ ‘ قرار گیرد.

فرمان INFORMATION یا INFO به شما کمک می‌کند تا از آنچه در یک فایل بازیافته قرار دارد اطلاع پیدا کنید :

MTB > INFO

ذخیره‌سازی یک فایل خروجی Minitab

از فرمان OUTFILE می‌توان در ذخیره‌سازی فایل خروجی که شامل پاسخ‌ها و تجزیه و تحلیل داده‌هاست استفاده نمود. به‌طور مثال شما یک سری عملیات آماری روی داده‌های مثال ۲ انجام داده‌اید و می‌خواهید فایل خروجی با نام SCORE را ذخیره کنید. به‌یاد داشته باشید نام فایل

بین ' ' قرار می‌گیرد و این کار باید قبل از آن که هرگونه عملیات آماری انجام شود، پس از آمدن اولین PROMPT در MINITAB انجام شود :

این فرمان فایل خروجی را روی دیسک سخت ذخیره می‌کند.

MTB > OUTFILE 'SCORE'

این فرمان فایل خروجی را روی Floppy در درایو A ذخیره می‌سازد.

MTB > OUTFILE 'A: SCORE'

اگر می‌خواهید فایل خروجی در دایرکتوری خاصی چون STAT ذخیره شود چنین عمل کنید :

MTB > OUTFILE 'C: \ STAT \ SCORE'

پس از آن که کارتان با MINITAB تمام شد، از محیط MINITAB خارج شده و فایل SCORE.LIS را چاپ کنید یا در جای دیگر که لازم دارید کپی کنید. با اجرای دستور OUTFILE، پسوند LIS . بطور اتوماتیک توسط MINITAB به‌نام فایل اتصال پیدا می‌کند. اگر پسوند خاصی را درنظر دارید می‌توانید از آن نام ببرید :

OUTFILE 'SCORE. TXT'

یا

OUTFILE 'A: SCORE. TXT'

به این ترتیب فایل خروجی با نام SCORE. TXT ذخیره می‌شود.

چاپ فایل خروجی MINITAB

شما مستقیماً در محیط MINITAB نمی‌توانید یک فایل خروجی چون SCORE.LIS یا SCORE.TXT را چاپ کنید برای این کار یا باید ابتدا به محیط MINITAB FOR WINDOW رفته و همان‌طور که قبلاً گفته شد از FILE در لیست انتخاب استفاده کنید و فایل خروجی را همان‌طور که قبلاً توضیح داده شده است چاپ کنید. روش دیگر این است که ابتدا از محیط MINITAB خارج شوید و سپس فایل را چاپ کنید.

ورود داده‌ها با استفاده از فرمان READ

اگر از فرمان‌های زبانی MINITAB استفاده می‌کنید و دو یا بیش از دو متغیر کمّی با تعداد مساوی داده برای هریک دارید، می‌توانید از READ استفاده کنید مثال ۴ نشان‌دهنده چگونگی این کار است.

مثال ۴: فرض کنید می‌خواهید داده‌های روبرو را برای قد و وزن شش نفر وارد کنید:

برای ورود این داده‌ها با استفاده از فرمان READ قد و وزن هر فرد را در یک ردیف وارد کنید.

قد (سانتی‌متر)	وزن (کیلو)
۱۷۲	۸۱
۱۶۷	۶۱
۱۶۳	۵۵
۱۷۸	۹۵
۱۷۰	۶۸
۱۶۵	۶۵

```
MTB > READ C1   C2 *
DATA >  172     81
DATA >  167     61
DATA >  163     55
DATA >  178     95
DATA >  170     68
DATA >  165     65
DATA > END
```

* این فرمان به MINITAB اعلام می‌دارد که شما می‌خواهید مقادیر دو متغیر را در دو ستون C۱ و C۲ وارد کنید.

انتخاب نمونه با استفاده از MINITAB

فرض کنید داده‌های ورودی شما در مثال ۲ متعلق به یک جامعهٔ ۸ نفری باشد و شما می‌خواهید مشاهده‌ای روی یک نمونهٔ ۳ نفری داشته باشید و داده‌های این نمونه را در ستون‌های C۳ و C۴ ذخیره کنید. دستورهای زیر چگونگی انجام این کار را نشان می‌دهند.

```
MTB > NOTE : SELECTING A SAMPLE OF OBSERVATIONS
MTB > SAMPLE 3 FROM C1 – C2 PUT IN C3 – C4
MTB > PRINT C3 – C4
```

این فرمان داده‌های نمونه را در ستون‌های C۳ و C۴ به نمایش می‌گذارد.

خروج از MINITAB

برای خروج از MINITAB زمانی که از فرمان‌های زبانی MINITAB استفاده می‌کنید، باید از دستور STOP پس از PROMPT در MINITAB استفاده نمود:

```
MTB > STOP
```

MINITAB از شما سؤالاتی می‌پرسد که باید به آن‌ها پاسخ بلی یا خیر دهید. پس از پاسخ دادن به تمام سؤالات از MINITAB خارج می‌شوید.

فرمان‌های دیگر در MINITAB

MTB > HELP HELP

از این فرمان می‌توان برای کمک گرفتن در MINITAB استفاده نمود.

MTB > HELP COMMANDS

این فرمان کمک در رابطه با فرمان‌های MINITAB ارائه می‌کند.

MTB > HELP OVERVIEW

از این فرمان می‌توان برای مرور MINITAB استفاده نمود.

MTB > COPY C1 TO C2

این فرمان مقادیر ستون C۱ را در ستون C۲ کپی می‌کند.

MTB > ERASE C2

این فرمان مقادیر موجود در ستون C۲ را پاک می‌کند.

MTB > DELETE ROW 2 C1

این فرمان ردیف دوم (مقدار دوم) از ستون C۱ را حذف می‌کند.

MTB > DELETE ROW 2 C1 – C2

این فرمان ردیف دوم در ستون‌های C۱ و C۲ را حذف می‌کند.

MTB > INSERT BETWEEN 2 and 3 C1 – C2

این فرمان ردیف جدید بین ردیف‌های دوم و سوم برای ستون‌های اول و دوم ایجاد می‌کند.

MTB > LET C1 (4) =10

این فرمان عدد ۱۰ را در چهارمین ورودی ستون اول وارد می‌کند.

MTB > SORT C1 PUT IN C3

این فرمان داده‌های ستون اول را به ترتیب صعودی مرتب کرده و در ستون سوم می‌نویسد.

MTB > ADD C1 C2 PUT IN C4

این فرمان مقادیر متناظر را در ستون‌های اول و دوم باهم جمع می‌کند و در ستون چهارم قرار می‌دهد.

MTB > SUBTRACT C2 FROM C1 PUT IN C5

این فرمان مقدار C۲ را از مقدار C۱ متناظر کم کرده و در ستون پنجم قرار می‌دهد.

MTB > MULTIPLY C1 BY C2 PUT IN C6

این فرمان مقادیر متناظر C۱ و C۲ را درهم ضرب کرده و در C۶ قرار می‌دهد.

MTB > DIVIDE C1 BY C2 PUT IN C7

این فرمان هر مقدار موجود در C۱ را بر دادهٔ متناظر در C۲ تقسیم می‌کند، نتیجه را در C۷ (ستون هفتم) قرار می‌دهد.

MTB > LET C8 = C1* C2

این فرمان مقادیر C۱ و C۲ را درهم ضرب می‌کند و در C۸ (ستون هشتم) قرار می‌دهد.

MTB > LET C9 = C1** 2

این فرمان مقادیر C۱ را به توان ۲ می‌رساند و در C۹ (ستون نهم) قرار می‌دهد.

MTB > ADD 5 TO C1 PUT IN C10

این فرمان به هریک از مقادیر C۱ عدد پنج را اضافه کرده و در C۱۰ (ستون دهم) قرار می‌دهد.

MTB > SUBTRACT 8 FROM C1 PUT IN C11

این فرمان از هریک از مقادیر C۱ مقدار ۸ را کم کرده و در C۱۱ (ستون یازدهم) قرار می‌دهد.

MTB > MULTIPLY C1 BY 2 PUT IN C12

این فرمان مقادیر C۱ را در عدد ۲ ضرب کرده و در C۱۲ (ستون دوازدهم) قرار می‌دهد.

MTB > DIVIDE C1 BY 3 PUT IN C13

این فرمان مقادیر C۱ را بر ۳ تقسیم کرده و در C۱۳ (ستون سیزدهم) قرار می‌دهد.

فرمان HELP COMMAND توضیحات راجع به فرامین را نیز به همراه دارد.

به‌طور مثال HELP SET ، اطلاعاتی راجع به فرمان SET و موارد استفادهٔ آن به شما می‌دهد.

تمرین

۱ــ نام و قد و وزن همکلاسی‌های خود را روی یک دیسک Floppy ثبت کنید. سپس یک نمونه ۵ نفری با استفاده از فرمان‌های زبانی MINITAB و نیز MINITAB WINDOWS انتخاب کرده و داده‌های مربوط به این نمونه را چاپ کنید.

۲ــ جدول زیر نام ساعات کار و مزد پنج کارمند را نشان می‌دهد.

نام	ساعات کار	مزد (تومان)
علی	۴۲	۱۷۲۵۰
شیوا	۳۳	۱۴۸۳۰
زهرا	۲۸	۱۳۵۵۰
بهمن	۴۷	۱۷۹۰۰
اکبر	۴۰	۱۸۲۰۰

الف ــ با استفاده از MINITAB WINDOW داده‌های فوق را ثبت کنید. فایل فوق را با نام WORKER. MTW ذخیره کنید. از MINITAB خارج شوید. سپس دوباره وارد MINITAB شوید و فایل WORKER. MTW را دوباره باز کنید.

ب ــ داده‌های فوق را در ستون‌های C۱، C۲ و C۳ با استفاده از فرمان‌های زبانی MINITAB وارد کنید. فایل فوق را با نام WORKER.MTW ذخیره کنید. فرمان‌های زیر را امتحان کنید و نتیجه را بررسی کنید :

```
MTW > PRINT C1 - C3
MTB > NAME C1 ‘NAME’ C2 ‘HOURS’ C3 ‘SALARY’
MTB > PRINT. ‘NAME’ ‘HOURS’ ‘SALARY’
MTB > LET K2 = SUM (C2)
MTB > PRINT K2
MTB > LET K3 = SUM (C3)
MTB > PRINT K3
MTB > PRINT C4
MTB > LET K4 = SUM (C4)
MTB > PRINT K4
MTB > LET C5 = C3 * C3
MTB > LET K5 = SUM (C5)
MTB > PRINT C2 - C5
MTB > PRINT K2 - K5
MTB > SAMPLE 2 FROM C1 - C3 PUT IN C6-C8
MTB > PRINT C6 - C8
MTB > SAVE ‘ASSIGN1’
MTB > DIR
MTB > RETRIEVE ‘ASSIGN1’
MTB > INFO
MTB > STOP
```


هیچ کس چیزی را در دل نهان
نکرد، جز که در سخنان
بی‌اندیشه‌اش آشکار گشت و
در صفحهٔ رخسارش پدیدار.

فصل پنجم

نمودارها و تحلیل داده‌ها

فکر می‌کنید کسب اطلاعات با استفاده از جدول آسان‌تر است یا نمودار و یا یک متن؟ شنیده‌ایم که می‌گویند : «یک تصویر خوب ارزش هزار کلمه را دارد.» با دیدن یک نمودار در یک نگاه می‌توان به بسیاری از ویژگی‌های مجموعه‌ای از اطلاعات پی‌برد. امروزه ارائهٔ نموداری داده‌های آماری بسیار رایج است.

پوسترها، روزنامه‌ها، مجلات، تلویزیون از نمودار برای به‌تصویردرآوردن مسائل و ارائهٔ آن‌ها به طریقی ملموس‌تر استفاده می‌کنند.

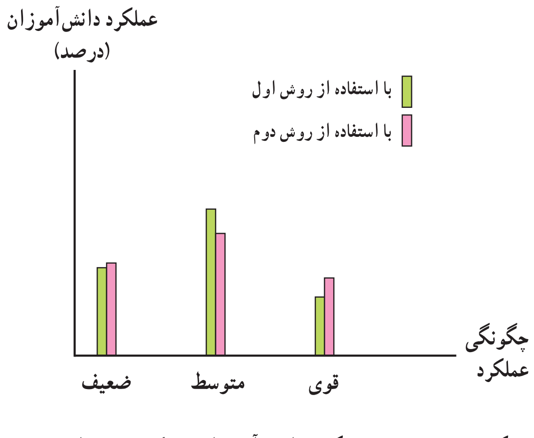
اگر به‌خاطر داشته باشید در فصل اول از شما پرسیدیم که روزانه با چه کسانی سر و کار دارید؟ سن و سال آن‌ها چه‌قدر است؟ تا چه سطحی تحصیل کرده‌اند؟ یک جواب ساده به این سؤال می‌توانست به یکی از صورت‌های زیر باشد :

ــ با تعداد کمی از افراد که اکثر آن‌ها جوان و تحصیل‌کرده هستند.

ــ با تعداد زیادی از افراد که اکثر آن‌ها مسن و با تحصیلات پایین هستند.

ــ با تعداد زیادی از افراد که بیش‌تر آن‌ها تا حد دیپلم درس خوانده‌اند و میان‌سال هستند.

این سه پاسخ در یک جمله کوتاه توانسته‌اند دورنمایی از جمعیتی را که روزانه شما با آن‌ها سر و کار دارید، ترسیم کنند. آیا سریع‌تر از بیان این جمله می‌توانستید همان جمعیت را معرفی کنید؟ یا در توفیق یکی از روش‌های آموزشی (مثالی که در فصل قبل بررسی شد) مطالعه جدول و توضیحات اگرچه ما را به اطلاعات عمیق‌تر می‌رساند ولی فعلاً لازم داریم در یک نگاه حقایق زیادی را دریابیم. یکی از این‌روش‌ها رسم شکل روبرو است.



شکل ۱ ــ مقایسهٔ عملکرد دانش‌آموزان دو کلاس در امتحان

نمودارها یا **شاخص‌های هندسی** وسیله‌ای سودمند برای به‌تصویردرآوردن و تجسم جامعه می‌باشند. در این فصل به معرفی برخی از این شاخص‌های هندسی می‌پردازیم.

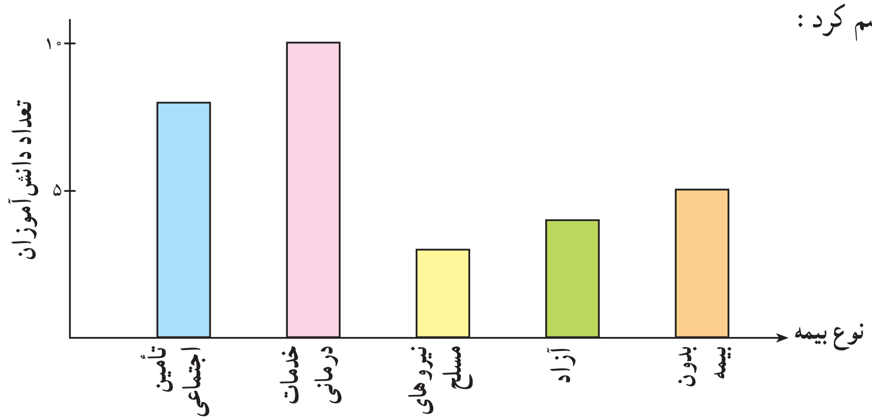
نمودار میله‌ای

جدول ۱ نشان‌دهندهٔ پاسخ‌های دانش‌آموزان کلاس دربارهٔ نوع بیمهٔ درمانی مورد استفاده آن‌هاست.

جدول ۱ــ وضعیت بیمهٔ دانش‌آموزان کلاس

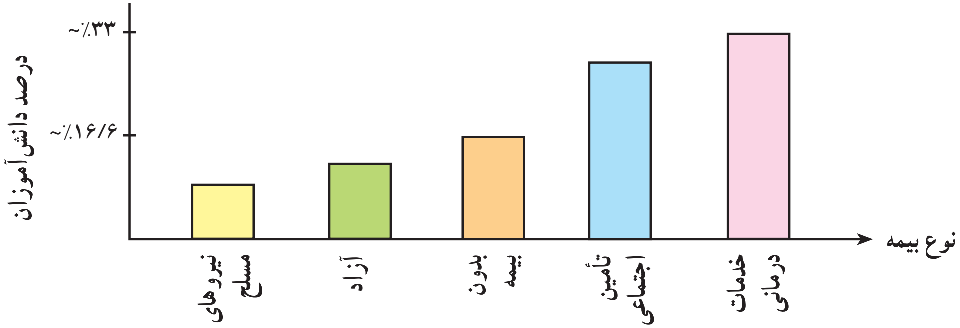
بیمه تأمین‌اجتماعی	۸
بیمه خدمات درمانی	۱۰
بیمه نیروهای مسلح	۳
شرکت‌های بیمهٔ آزاد	۴
دانش‌آموزانی که بیمه ندارند	۵

این اطلاعات را می‌توان در شکل زیر مجسم کرد :



شکل ۲ ــ نمودار توزیع دارندگان انواع بیمه در یک کلاس

این نمودار بیش‌تر برای متغیرهای گسسته و کیفی مناسب است. ترتیب قرارگرفتن میله‌ها اهمیت ندارد. آن‌چه که در این نمودار مهم است مقایسه فراوانی انواع بیمه است. از این‌رو اگر این نمودار را با درصد رسم کنیم، بهتر خواهد بود و هم‌چنین اگر به گونه‌ای رسم کنیم که فراوانی‌ها از بزرگ به کوچک و یا از کوچک به بزرگ رسم شوند، عمل مقایسه سریع‌تر انجام می‌شود.



شکل ۳ــ نمودار توزیع دارندگان انواع بیمه در یک کلاس

کم‌بودن درصد استفاده‌کنندگان از بیمه نیروهای مسلح به این علت نیست که خدمات قابل قبولی ارائه نمی‌کند، بلکه به این علت است که افراد خاصی (نظامی) می‌توانند از این بیمه بهره بگیرند. هم‌چنین زیادبودن درصد استفاده‌کنندگان از خدمات درمانی هم نباید به این نتیجه منتهی شود که خدمات آن مورد قبول افراد بیش‌تری است، این زیادبودن به دلیل جمعیت کثیری است که از این بیمه استفاده می‌کنند.

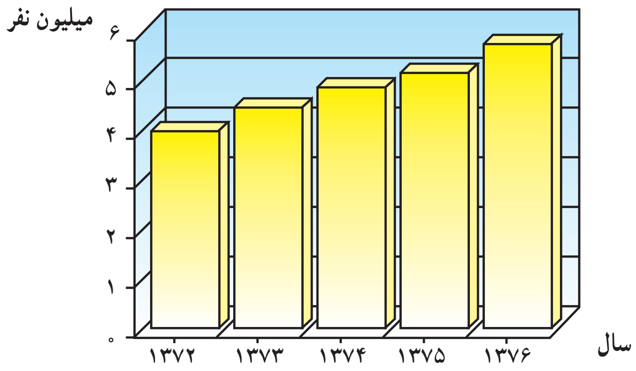
نمودار میله‌ای شامل قسمت‌های زیر است :

عنوان: زیر هر نمودار باید موضوع مورد مطالعه به‌طور خلاصه نوشته شود. در نمودار صفحهٔ قبل عنوان نمودار عبارت است از : توزیع‌دارندگان انواع بیمه در کلاس.

برچسب محورها: هریک از محورها برچسبی دارند که مشخص‌کننده متغیری است که آن محور نشان می‌دهد. در شکل ۲، محور افقی مشخص‌کننده انواع بیمه‌ها و محور عمودی مشخص‌کننده تعداد استفاده‌کنندگان از این بیمه‌هاست.

مقیاس: مقیاس هر محور باید مشخص باشد. مثلاً در مثال بالا اگر بخواهیم اطلاعات مربوط به استفاده‌کنندگان از بیمه در سطح کشور را مطالعه کنیم در این صورت تعداد استفاده‌کننده برای برخی از بیمه‌ها از میلیون هم تجاوز می‌کند. پس مشخص‌کردن یک واحد برای هر فرد منجر به رسم میله‌های بسیار طولانی می‌گردد که در صفحه نخواهد گنجید. از این‌رو برای هر ۱۰۰ نفر یا هر ۱۰۰۰ نفر یک واحد انتخاب می‌شود که در این صورت باید در حاشیهٔ نمودار این نکته تذکر داده شود. در نمودار زیر تعداد بیمه‌شدگان تحت پوشش سازمان تأمین اجتماعی را در سال‌های مختلف ملاحظه می‌کنید.





شکل ۴ــ بیمه‌شدگان اصلی تحت پوشش سازمان تأمین اجتماعی

با توجه به نمودار بالا به سؤالات زیر پاسخ دهید :

۱ــ هر واحد روی نمودار نشان‌دهنده چند نفر می‌باشد؟

۲ــ در کدام سال نفرات بیش‌تری و در کدام سال نفرات کم‌تری تحت پوشش سازمان تأمین اجتماعی قرار گرفتند؟

۳ــ در کدام سال (ها) تعداد بیمه‌شدگان بین ۴ میلیون و ۵ میلیون است؟

۴ــ تقریباً چند نفر در سال ۱۳۷۵ بیش‌تر از سال ۱۳۷۳ زیر پوشش بیمه قرار گرفته‌اند؟

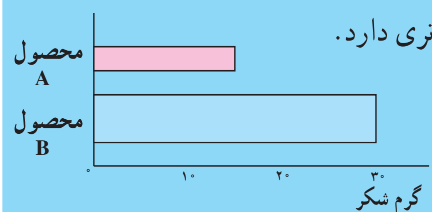
۵ــ آیا می‌توانید در مورد تعداد دارندگان بیمه در سال ۱۳۸۰ پیش‌بینی کنید؟

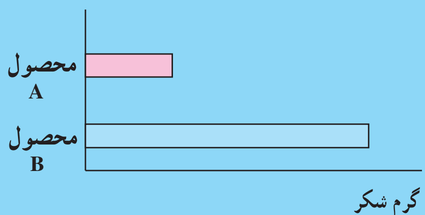
همان‌طوری‌که در نمودارها دیده‌اید برای رسم نمودار میله‌ای کافی است، متغیر تصادفی مورد مطالعه را روی محور xها مشخص کنیم و روی هر یک از آن‌ها میله‌ای رسم کنیم که طول میله **متناسب با فراوانی** آن مقدار باشد. طول میله‌ها بیان‌گر فراوانی و یا درصد فراوانی نسبی نظیر آن مقدار خواهد بود.

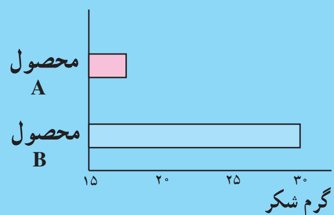
بحث کلاسی ۱

دو شرکت A و B نوعی شیرینی تولید می‌کنند. در نمودارهای زیر شرکت B سعی دارد اعلام کند شیرینی‌های او شکر بیش‌تری دارد.

ــ چرا این نمودارها گمراه‌کننده هستند؟







بحث کلاسی ۲

اطلاعات زیر را در نظر بگیرید :

پیش‌دانشگاهی	سوم دبیرستان	دوم دبیرستان	اول دبیرستان	سوم راهنمایی	دوم راهنمایی	اول راهنمایی	
۲۹۴	۵۱۵	۵۰۳	۶۰۳	۷۴۷	۸۰۶	۸۴۹	تعداد دختران بر حسب ۱۰۰۰ نفر

(برگرفته از آمار آموزش و پرورش، معاونت برنامه‌ریزی و نیروی انسانی)

جدول ۲

ــ برای رسم نمودار میله‌ای به چه اطلاعاتی نیاز دارید؟

ــ نمودار میله‌ای اطلاعات به‌دست‌آمده را با استفاده از فراوانی مطلق و فراوانی نسبی رسم کنید.

ــ این دو نمودار چه تفاوت‌هایی دارند؟

ــ آیا این تفاوت‌ها تأثیری در تعبیر داده‌ها دارند یا خیر؟ توضیح دهید.



نمودار مستطیلی

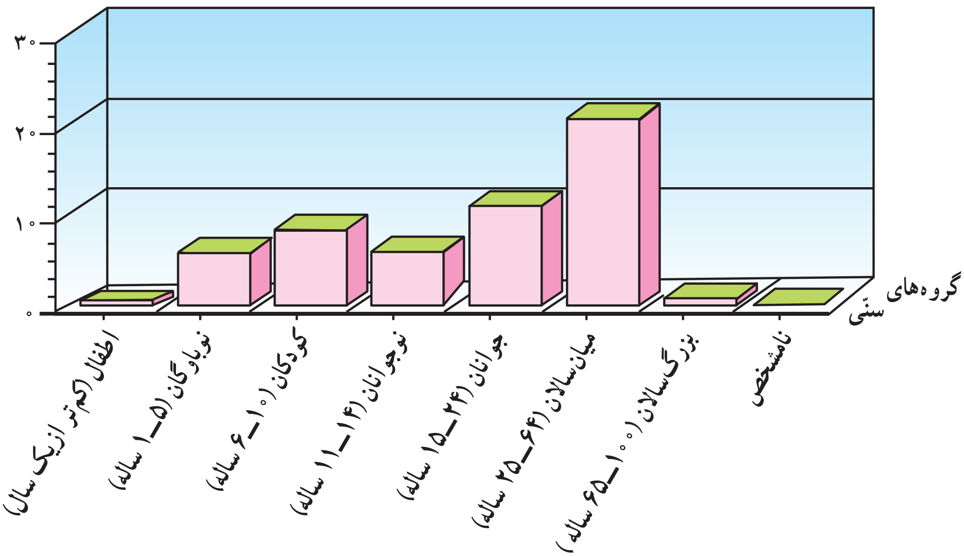
این نمودار برای متغیرهای کمی پیوسته مناسب است. در جدول زیر، اطلاعات مربوط به متغیر سن جمعیت ایران آمده است. می‌دانیم سن یک متغیر پیوسته است (در صورت لزوم می‌توانیم از واحدهای کوچک‌تر از قبیل روز و ساعت استفاده کنیم ولی از آن‌ها صرف‌نظر می‌شود) سن از نوع زمان است و لذا پیوسته است. ولی گاهی برحسب فاصله‌های سنی جمعیت را به دسته‌هایی که در جدول روبرو ملاحظه می‌کنید تقسیم می‌کنند. این دسته‌ها را گروه‌های سنی می‌نامند.

جدول ۳

گروه عمده سنی	مرد و زن
جمع	۶۰۰۵۵۴۸۸
اطفال (کم‌تر از یک‌سال)	۱۰۲۰۹۳۶
نوباوگان (۵ــ۱ ساله)	۶۷۳۲۷۸۷
کودکان (۱۰ــ۶ ساله)	۸۷۶۹۷۳۷
نوجوانان (۱۴ــ۱۱ ساله)	۷۲۰۲۰۸۵
جوانان (۲۴ــ۱۵ ساله)	۱۲۳۳۷۵۲۹
میان‌سالان (۶۴ــ۲۵ ساله)	۲۱۳۶۴۸۷۷
بزرگ‌سالان (۱۰۰ــ۶۵ ساله)	۲۵۹۵۱۸۱
نامشخص	۳۲۳۵۶

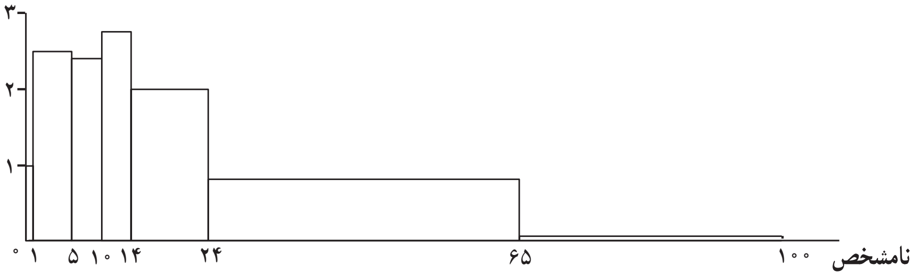
مأخذ: مرکز آمار ایران.

نمودار مربوط به این اطلاعات را ممکن است به‌صورت نمودار میله‌ای به‌شکل زیر رسم کنیم.



شکل ۵ ــ نمودار توزیع جمعیّت در گروه‌های مختلف سنّی

از آن‌جایی که سن، یک متغیر پیوسته است، انتظار داریم این پیوستگی به‌طریقی در نمودار دیده شود از این‌رو از نمودار مستطیلی استفاده می‌کنیم. در این نمودار مستطیل‌هایی رسم می‌کنیم که قاعدهٔ آن‌ها روی محور xها و برابر طول هر یک از دسته‌هاست و ارتفاع آن‌ها به موازات محور yها و متناسب با فراوانی دسته‌هاست. مثلاً نمودار مستطیلی مربوط به داده‌های جدول بالا به‌صورت زیر است : (۵ ــ ۱ یعنی تا ۶ سال، پنج سال تمام)



شکل ۶ ــ نمودار توزیع جمعیّت

تذکر: در نمودار مستطیلی مساحت مستطیل‌ها با یک‌دیگر مقایسه می‌شوند هرچقدر مساحت مستطیل بیشتر باشد تعداد کسانی که در آن دسته قرار می‌گیرند بیشتر است و در صورتی که طول دسته‌ها (قاعدهٔ مستطیل‌ها) با یک‌دیگر برابر باشند، ارتفاع مستطیل‌ها (فراوانی دسته‌ها) با یک‌دیگر مقایسه خواهند شد.

در این کتاب برای سهولت مقایسه‌ها، طول دسته‌ها را مساوی در نظر خواهیم گرفت.

نمودار مستطیلی نمایشی از داده‌های دسته‌بندی شده است که در آن سطح مستطیل‌ها متناسب با فراوانی دسته‌ها است.

بحث کلاسی

با توجه به نمودار جمعیت (شکل ۶) به سؤالات زیر پاسخ دهید.

ــ در یک نمودار مستطیلی، طول قاعدهٔ ستون‌ها نشان‌گر چیست؟ ارتفاع ستون‌ها نشان‌گر چیست؟

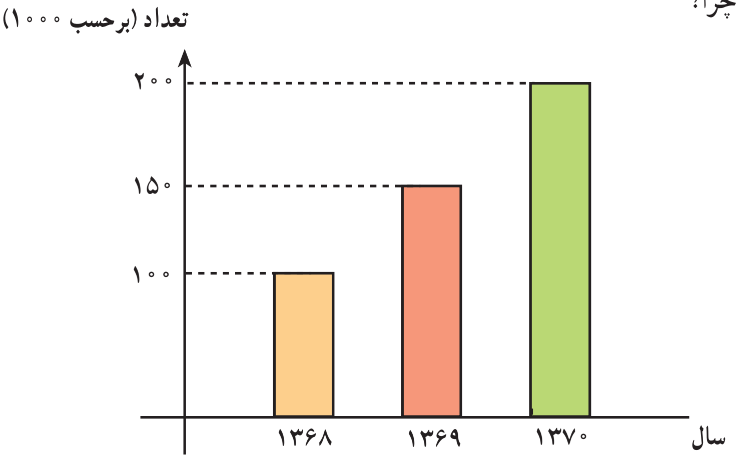
ــ اگر به‌طور متوسط دانش‌آموزان در سن ۱۵ سالگی در کلاس اول دبیرستان مشغول به تحصیل باشند، آیا می‌توانید با استفاده از نمودار فوق بگویید در این سال چند دانش‌آموز در کلاس اول دبیرستان مشغول به تحصیل هستند؟

ــ با استفاده از نمودار، جمعیت جوان و نوجوان کشور را تخمین بزنید.

ــ همان‌طور که مشاهده می‌شود با استفاده از نموداری چون نمودار ستونی می‌توان به سؤال‌های مربوط به مجموعه‌ای از داده‌ها سریع‌تر پاسخ داد. اگر شما دست‌اندرکار طراحی و برنامه‌ریزی در امر آموزش باشید، با توجه به نمودار فوق، (شکل ۶) سؤال‌هایی طرح کنید و به آن‌ها در کلاس پاسخ دهید.

تمرین

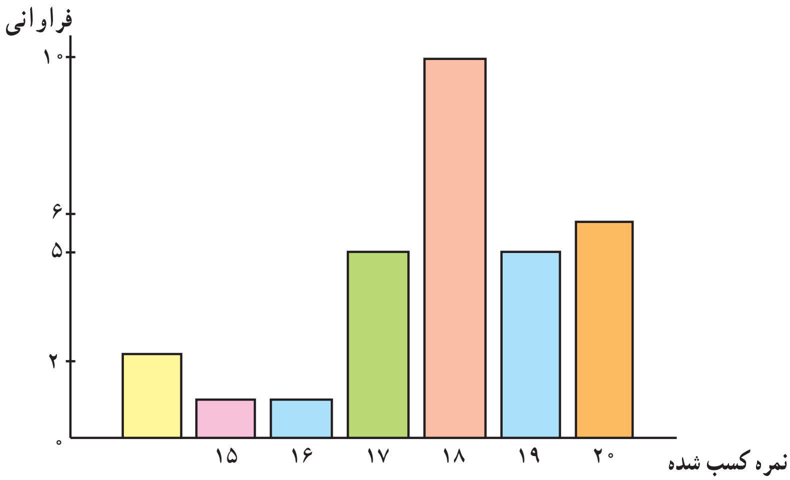
۱ــ نمودار زیر قسمتی از نمودار تولید یک کارخانه را نشان می‌دهد. آیا این نمودار نشان‌دهندهٔ افزایش تولید در سال‌های ۱۳۶۸ تا ۱۳۷۰ است؟ این نمودار به همین شکل که هست، ما را دچار اشتباه می‌کند. چرا؟

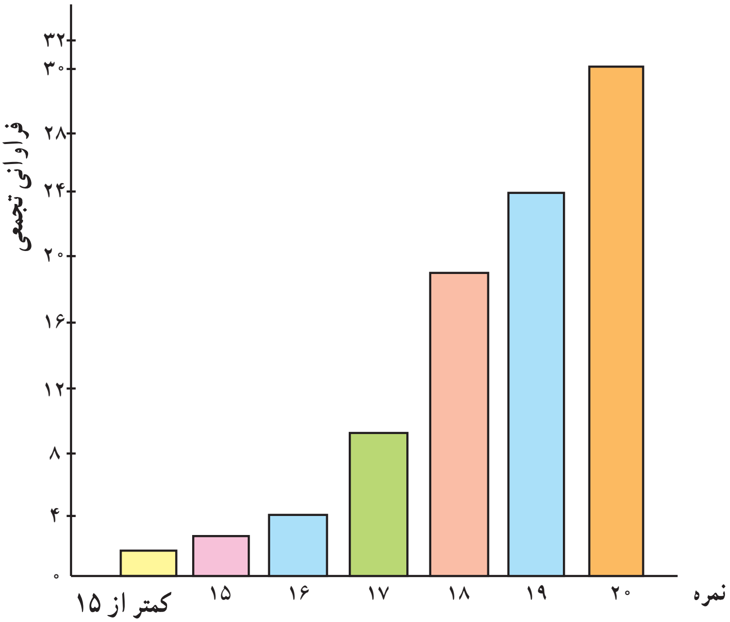


۲ــ در یک امتحان کتبی از درس کمک‌های اولیه، افرادی که نمره بیش‌تر از ۱۵ (از ۲۰) کسب کرده باشند، می‌توانند دورهٔ عملی را شروع کنند. نمودارهای زیر نتیجهٔ آزمون را نشان می‌دهد. (نمرات به شکل عدد صحیح ثبت شده‌اند.)

ــ با توجه به نمودارها تعداد افرادی را که می‌توانند در دوره شرکت نمایند مشخص کنید.

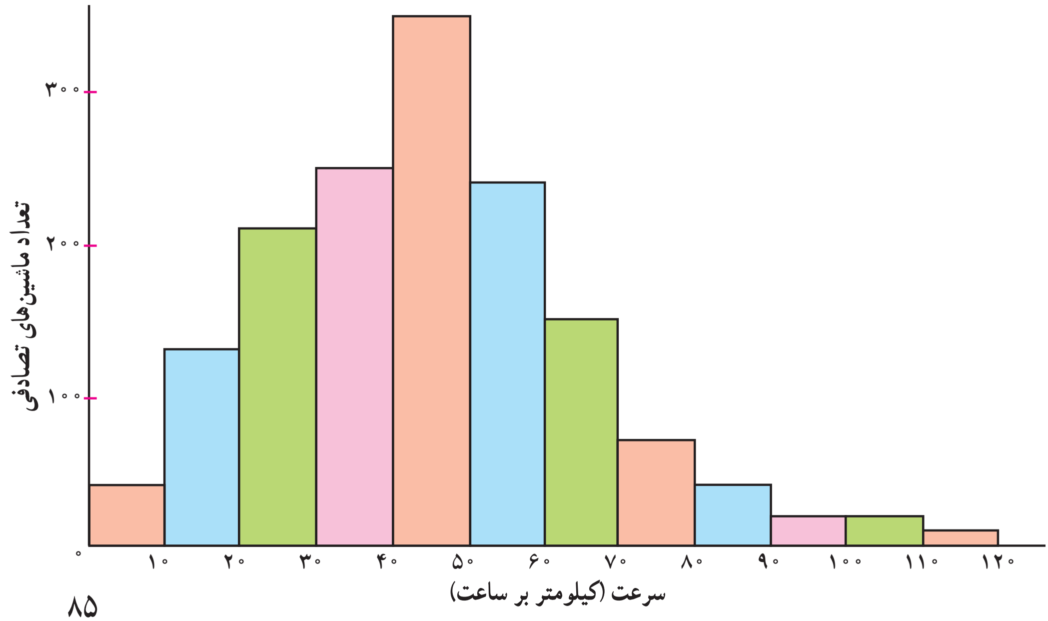
ــ کدام نمودار شما را سریع‌تر به پاسخ هدایت می‌کند.





۳ــ جدول زیر نشان‌دهندهٔ تعداد تصادف‌های اتومبیل‌ها و سرعت حرکت آن‌ها در زمان تصادف است.

سرعت (کیلومتر / ساعت)	۰—۱۰	۱۰—۲۰	۲۰—۳۰	۳۰—۴۰	۴۰—۵۰	۵۰—۶۰	۶۰—۷۰	۷۰—۸۰	۸۰—۹۰	۹۰—۱۰۰	۱۰۰—۱۱۰	۱۱۰—۱۲۰
تعداد ماشین‌های تصادف‌کرده	۴۰	۱۳۰	۲۱۰	۲۵۰	۳۵۰	۲۴۰	۱۵۰	۷۰	۴۰	۲۰	۲۰	۱۰



همان‌طور که دیده می‌شود تعداد تصادف‌های اتومبیل‌هایی که با سرعت بالاتر از ۹۰ کیلومتر در ساعت رفته‌اند کم‌تر است. پس :

«هرچه سریع‌تر بروید، مطمئن‌تر و امن‌تر است»

نتیجه‌گیری فوق چه اشکالی دارد؟ چرا این تصور ایجاد شده است؟ برای تحلیل صحیح، چه نموداری باید رسم کرد؟

۴ــ ۵۰ دانش‌آموز یک کلاس در آزمون سنجش IQ (بهرهٔ هوشی) شرکت نمودند. نتایج به‌دست‌آمده به‌شرح زیر است :

۱۲۲ , ۱۱۷ , ۱۱۴ , ۱۱۲ , ۱۳۰ , ۹۲ , ۱۱۸ , ۱۰۹ , ۱۱۳ , ۱۱۵
۱۱۸ , ۱۱۶ , ۱۰۶ , ۱۱۱ , ۱۱۲ , ۱۲۴ , ۱۱۳ , ۱۰۸ , ۱۰۷ , ۱۲۷
۱۱۴ , ۱۰۰ , ۱۰۳ , ۱۱۵ , ۱۲۴ , ۱۱۰ , ۱۱۸ , ۱۱۸ , ۱۰۷ , ۱۲۱
۱۱۰ , ۱۱۶ , ۱۲۶ , ۱۲۱ , ۱۳۵ , ۱۰۴ , ۱۲۳ , ۱۱۶ , ۱۲۴ , ۱۰۴
۱۱۲ , ۱۱۷ , ۱۱۳ , ۱۰۷ , ۱۰۳ , ۱۰۲ , ۱۲۹ , ۹۸ , ۱۳۴ , ۹۶

ــ جدول فراوانی را برای این داده‌ها تشکیل دهید. برای این جدول ۹ گروه با طول دستهٔ مساوی در نظر بگیرید.

ــ نمودار مستطیلی این داده‌ها را رسم کنید.

ــ چه نتیجه‌ای می‌گیرید.

۵ــ با توجه به اطلاعات داده شده در جدول صفحهٔ بعد :

ــ برای رسم نمودار میله‌ای برای اطلاعات داده شده چه مشکلاتی خواهید داشت؟

ــ برای رسم نمودار مستطیلی به چه اطلاعاتی نیاز دارید؟

ــ نمودار مستطیلی مربوط به فضاهای سبز مراکز استان‌های ایران را با در نظر گرفتن ۴ دسته رسم کنید.

ــ نمودار مستطیلی رسم‌شده چه اطلاعاتی در اختیار شما قرار می‌دهد؟

مساحت بوستان‌های طبیعی، فضاهای سبز و ذخیره‌گاه‌های جنگلی تا پایان سال ۱۳۷۱ (هکتار)

استان	جمع کل عرصه‌های جنگلی	بوستان‌های طبیعی	فضاهای سبز	ذخیره‌گاه‌های جنگلی
جمع	۲۹۶۲۲۸	۳۳۸۴۱	۲۳۳۹۰۴	۲۸۴۸۳
تهران	۱۲۵۳۰	۰	۱۲۵۳۰	۰
مرکزی	۸۵۰	۰	۸۵۰	۰
گیلان	۳۴۳۵۶	۰	۳۳۳۵۶	۱۰۰۰
مازندران	۱۲۹۸۰۳	۸۶۳۱	۱۱۲۴۲۷	۸۷۴۵
آذربایجان شرقی	۱۰۶۰	۱۰۰	۸۶۰	۱۰۰
آذربایجان غربی	۶۰۹	۰	۶۰۹	۰
کرمانشاه	۱۲۶۳	۰	۱۲۳۳	۳۰
خوزستان	۱۵۹۱۰	۱۲۰۰۰	۳۸۶۷	۴۳
فارس	۱۱۲۲۳	۹۱۶۵	۱۹۰۸	۱۵۰
کرمان	۱۶۸۹	۰	۱۳۸۹	۳۰۰
خراسان	۱۰۳۱۳	۰	۳۱۳	۱۰۰۰۰
اصفهان	۳۳۴۰	۰	۳۳۲۰	۲۰
سیستان و بلوچستان	۲۷۶۳۷	۰	۲۷۵۸۷	۵۰
کردستان	۲۸۳۴	۱۰۰۰	۱۵۳۴	۳۰۰
همدان	۱۳۷۴	۰	۱۱۷۴	۲۰۰
چهارمحال‌وبختیاری	۲۶۸۲	۶۵۰	۱۳۴۷	۶۸۵
لرستان	۳۱۸۸	۸۵۰	۱۵۸۸	۷۵۰
ایلام	۳۵۰۱	۱۴۰۰	۲۱۰۱	۰
کهگیلویه و بویراحمد	۲۱۰۸	۴۵	۱۰۶۳	۱۰۰۰
بوشهر	۳۴۰۱	۰	۱۸۹۱	۱۵۱۰
زنجان	۳۹۶	۰	۳۹۶	۰
سمنان	۱۶۶۲	۰	۱۰۶۲	۶۰۰
یزد	۵۳۲	۰	۵۳۲	۰
هرمزگان	۲۳۹۶۷	۰	۲۰۹۶۷	۳۰۰۰

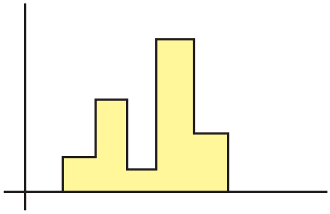
نمودار چندبر فراوانی

مجدداً به داده‌های مربوط به گروه سنی باز می‌گردیم. در رسم نمودار مستطیلی گفتیم که چون سن یک متغیر پیوسته است باید این پیوستگی را در نمودار آن نیز ببینیم. اما در نمودار مستطیلی فراوانی‌ها روی دسته‌ها تغییر نمی‌کنند و نمودار مستطیلی بیش‌تر شبیه به پلکانی به‌صورت روبرو است.

اگر بخواهیم تغییرات متغیر را بهتر نشان دهیم از نمودار چندبر فراوانی استفاده می‌کنیم.

داده‌هایی که در فصل سوم مربوط به تعداد کالاهایی که پس از فروش برای تعمیرات اولیه به کارخانهٔ سازنده عودت داده شده‌اند، در نظر بگیرید.

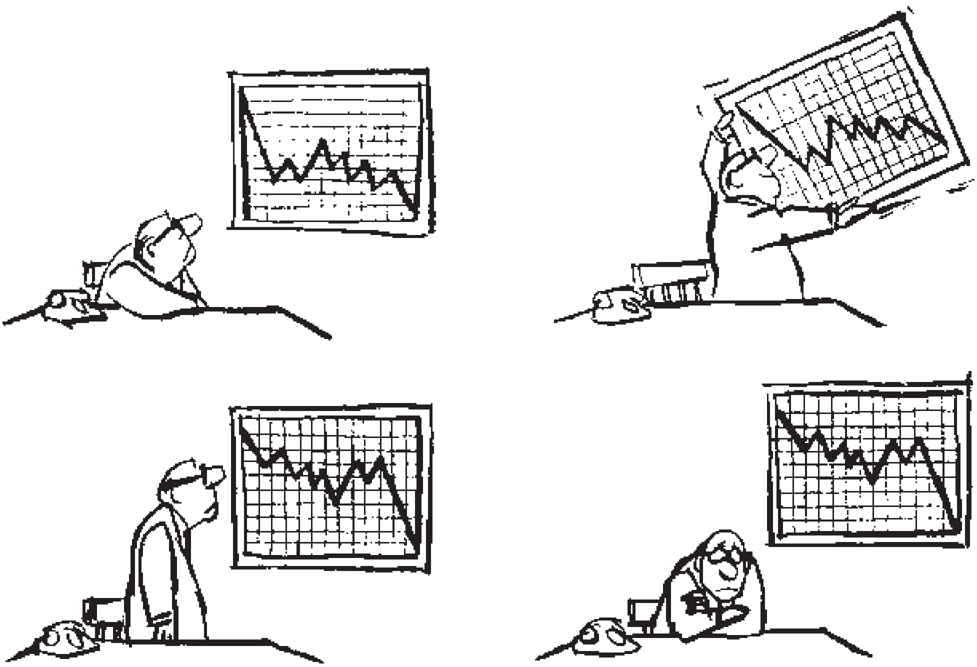
برای مثال فرض کنید بخواهیم یک نمودار چندبر فراوانی برای جدول فراوانی روبرو رسم کنیم.



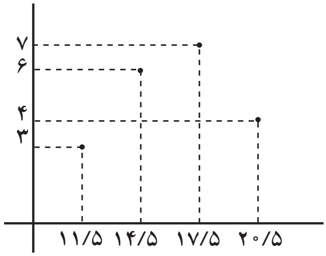
شکل ۸

جدول ۴

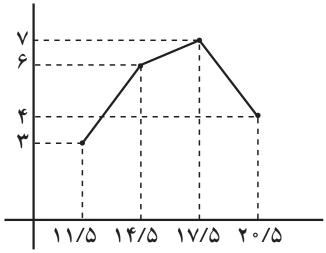
فراوانی	مرکز دسته‌ها	دسته‌ها
۳	۱۱/۵	۱۰ـ۱۳
۶	۱۴/۵	۱۳ـ۱۶
۷	۱۷/۵	۱۶ـ۱۹
۴	۲۰/۵	۱۹ـ۲۲
۲۰	جمع	



شکل ۷



شکل ۹



شکل ۱۰

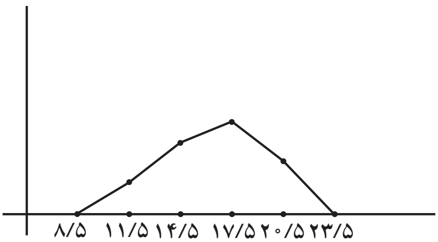
ابتدا نقاط (۱۱/۵،۳)، (۱۴/۵،۶)، (۱۷/۵،۷) و (۲۰/۵،۴) را در صفحهٔ محورهای مختصات رسم می‌کنیم. توجه دارید که طول این نقاط مرکز دسته‌ها و عرض آن‌ها فراوانی همان دسته است.

حال نقاط مذکور را متوالیاً به‌هم وصل می‌کنیم تا شکل ۱۰ حاصل شود.

برای آن‌که سطح زیر این منحنی برابر مساحت نمودار مستطیلی شود (این نکته بعدها مورد استفاده زیاد قرار می‌گیرد) دو سر نمودار فوق را به محور xها به‌طریقی که توزیع داده می‌شود وصل می‌کنیم. فرض کنیم دو دستهٔ اضافی یکی قبل از دستهٔ اول و یکی بعد از دسته آخر داشته باشیم، مسلماً فراوانی این دسته‌ها صفر است. پس نقاط نظیر این دسته‌ها عبارت خواهد بود از :

(۲۳/۵،۰) و (۸/۵،۰)

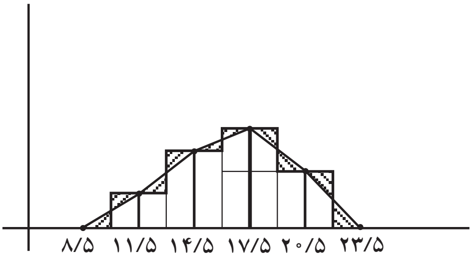
(دقت کنید ۳ = ۱۱/۵ − ۱۴/۵ = طول دسته، پس ۸/۵ = ۳ − ۱۱/۵ مرکز دسته فرضی ماقبل دستهٔ اول و ۲۳/۵ = ۳ + ۲۰/۵ مرکز دستهٔ فرضی آخر است) حال اگر این دو نقطه را به مجموعهٔ نقاط قبلی اضافه کنیم نمودار چندبر فراوانی به‌صورت زیر تکمیل می‌شود.



شکل ۱۱

برای جدول ۴ هر سه نمودار میله‌ای، مستطیلی و چندبر فراوانی را روی یک دستگاه محورها رسم می‌کنیم تا ارتباط آن‌ها مشخص‌تر شود.

از روی شکل ۱۲ می‌توانید توضیح دهید که چرا مساحت سطح زیر چندبر فراوانی و مساحت نمودار مستطیلی با هم برابر است؟ (به مساحت‌هایی که هاشور خورده است توجه کنید.)

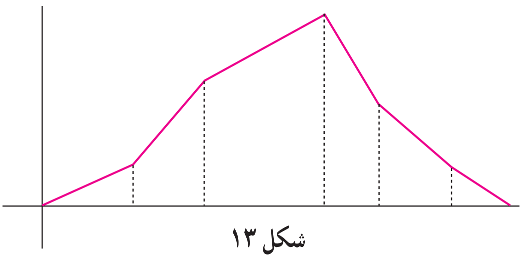


شکل ۱۲

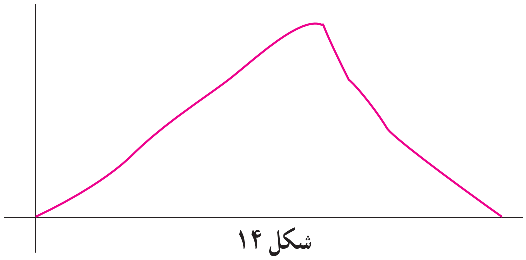
اگر به جای نمودار فراوانی،نمودار فراوانی نسبی را رسم کنیم، اطلاعات منسجم‌تری در اختیار

ما قرار می‌گیرد چون می‌توان فراوانی را با کل جامعه مقایسه کرد. نمودار چندبر فراوانی را می‌توان با داشتن فراوانی نسبی نیز رسم کرد و آن را چندبر فراوانی نسبی نامید.

چندبر فراوانی نسبی برای داده‌های پیوسته مناسب است و می‌دانیم که ما بر اساس داده‌های حاصل از یک نمونهٔ n تایی از یک جامعهٔ آماری به نمودار چندبر فراوانی دست یافتیم. حال ممکن است سؤال شود که اگر ما اندازهٔ نمونه را افزایش دهیم چه اتفاقی خواهد افتاد؟ چون تعداد داده‌ها زیاد می‌شود. بنابر توصیه‌هایی که در جدول فراوانی به‌عمل آمد، تعداد دسته‌ها افزایش خواهد یافت. و در این صورت معمولاً طول دسته‌ها کاهش خواهد یافت و در نتیجه چندبر فراوانی از پاره‌خط‌های بیش‌تری تشکیل خواهد شد. اگر به همین ترتیب ادامه داده و تعداد داده‌ها را افزایش دهیم در نهایت به یک چندبر فراوانی دست خواهیم یافت که بیش‌تر به یک منحنی هموار شبیه است تا یک چندبر فراوانی. در شکل زیر مراحل شکل‌گیری چنین منحنی را ملاحظه می‌کنید.



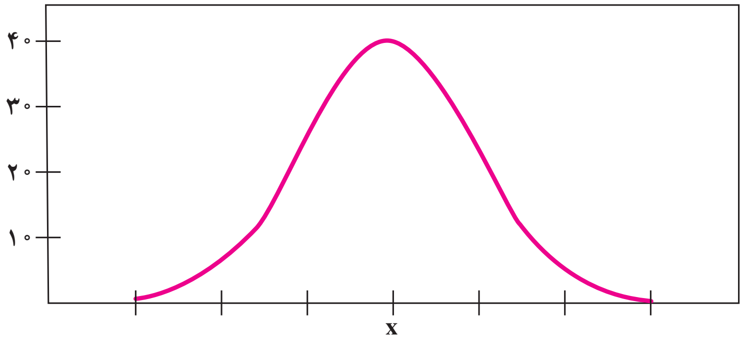
شکل ۱۳



شکل ۱۴

این منحنی هموار، بیان‌کننده وضعیت متغیر در جامعه است.

در اکثر مطالعات آماری این منحنی هموار شکل زیر را به خود می‌گیرد .



شکل ۱۵

این منحنی از منحنی‌های معروف در آمار است و در اکثر پدیده‌های طبیعی ظاهر می‌شود. ولی چون مقدمات کافی برای یادگیری آن را هنوز نمی‌دانیم به معرفی ظاهری آن پرداختیم. این منحنی متقارن و شبیه به زنگ است.

تمرین

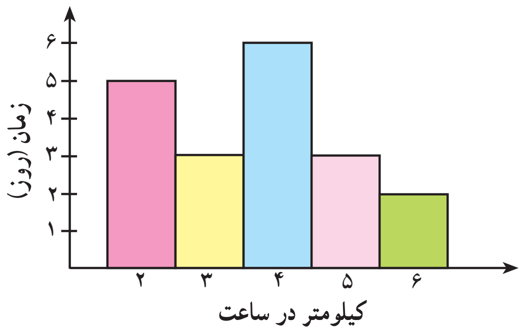
۱ ــ مدیر یک مؤسسه تاکسی‌سرویس مایل است بداند اتومبیل‌های این مؤسسه روزانه چند کیلومتر مسافرت می‌کنند. برای این منظور کارکرد ۲۴ اتومبیل را یادداشت کرد. داده‌های زیر به‌دست آمدند.

۱۱۸	۱۴۸	۱۲۸	۱۳۶	۱۰۷	۱۳۸
۱۲۳	۱۳۳	۱۲۳	۱۱۵	۱۲۹	۱۴۲
۱۳۵	۱۲۶	۱۴۴	۱۲۲	۱۲۸	۱۲۱
۱۲۶	۱۴۱	۱۵۳	۱۱۷	۹۸	۱۲۵

برای این داده‌ها یک جدول فراوانی تنظیم کنید. داده‌ها را در شش دسته، دسته‌بندی کنید. برای نمایش این داده‌ها چه نموداری مناسب است. نمودار چندبر فراوانی را رسم کنید.

۲ــ نمودار روبرو نشان‌دهندهٔ توزیع سرعت باد در ۱۹ روز متوالی می‌باشد. با استفاده از اطلاعات داده شده در نمودار، نمودار چندبر فراوانی را برای سرعت باد رسم کنید.

اگر حداقل سرعت باد لازم برای راندن یک قایق بادی ۵ کیلومتر در ساعت سرعت باد لازم باشد، چند روز برای راندن قایق بادی مناسب است؟



شکل ۱۶

چرا در این مسئله، نمودار چندبر فراوانی مناسب‌تر از نمودار مستطیلی است؟





نمودار دایره‌ای

یکی دیگر از نمودارهایی که می‌تواند اطلاعات موجود در داده‌ها را به‌سرعت در معرض دید قرار دهد **نمودار دایره‌ای** است. فرض کنید متغیر تصادفی مورد مطالعه ما دارای k حالت باشد (مثلاً در مطالعه مقطع تحصیلی، متغیر تصادفی کیفی دارای ۴ حالت ابتدایی، راهنمایی، دبیرستان و دانشگاه است) فرض کنید فراوانی حالت اول $f_۱$ و فراوانی حالت دوم $f_۲$ و الی آخر باشد، در این صورت دایره‌ای به شعاع دلخواه را به‌وسیله زاویه‌های مرکزی به k قسمت تقسیم می‌کنیم به‌قسمی که اندازهٔ زاویه مرکزی هریک از این قسمت‌ها متناسب با فراوانی آن قسمت باشد. پس زاویه مرکزی نظیر دستهٔ اول عبارت است از :

$$\frac{۳۶۰}{f_۱+\cdots+f_k}\times f_۱ = ۳۶۰\times\frac{f_۱}{n} \quad , \quad n = \text{تعداد کل داده‌ها}$$

و به همین ترتیب می‌توانیم سایر زوایا را به‌دست آوریم. هریک از این قسمت‌ها در دایره نمایش‌دهندهٔ سهم آن حالت از متغیر در جامعه است.

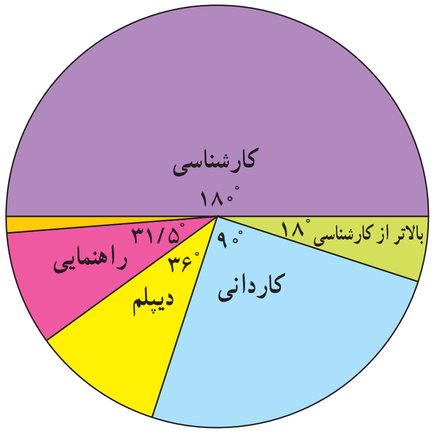
مثال: یک مؤسسه بزرگ ۴۰۰ نفر کارمند دارد که برحسب سطح سواد به‌صورت زیر توزیع شده‌اند.

سطح سواد	ابتدایی	راهنمایی	دیپلم	کاردانی	کارشناسی	بالاتر از کارشناسی
تعداد کارمندان	۵	۳۵	۴۰	۱۰۰	۲۰۰	۲۰

نمودار دایره‌ای مربوط به این داده‌ها به‌صورت زیر است :

زاویه‌ها مطابق دستور به‌شرح زیر محاسبه شده‌اند :

زاویه مرکزی برحسب درجه	فراوانی	حالت
$\frac{۳۶۰}{۴۰۰}\times ۵ = ۰/۹ \times ۵ = ۴/۵$	۵	ابتدایی
$۰/۹ \times ۳۵ = ۳۱/۵$	۳۵	راهنمایی
$۰/۹ \times ۴۰ = ۳۶$	۴۰	دیپلم
$۰/۹ \times ۱۰۰ = ۹۰$	۱۰۰	کاردانی
$۰/۹ \times ۲۰۰ = ۱۸۰$	۲۰۰	کارشناسی
$۰/۹ \times ۲۰ = ۱۸$	۲۰	بالاتر از کارشناسی

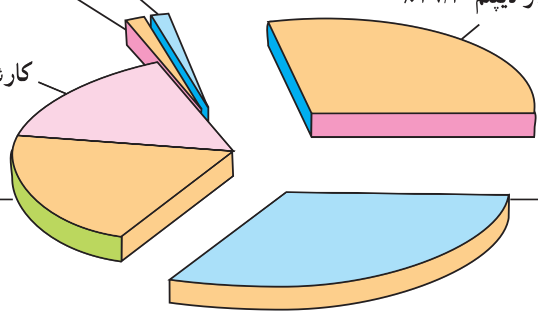


در شکل ۱۶ دیده می‌شود که ناحیه مربوط به کارشناسی بزرگتر از سایر نواحی است و برابر نصف دایره می‌باشد. ناحیه مربوط به ابتدایی از همه کم‌تر است. ناحیه کاردانی نصف ناحیه کارشناسی است. در هر صورت بزرگی و کوچکی این ناحیه‌ها (در مقایسه با یک‌دیگر) بیان‌کنندهٔ فراوانی بیش‌تر و یا کم‌تر برای آن حالت از متغیر است.

جدول زیر توزیع کارکنان تابع قانون کار را در سال ۱۳۷۶ برحسب مدرک تحصیلی نشان می‌دهد.

نوع مدرک	کمتر از دیپلم	دیپلم	کاردانی	کارشناسی	کارشناسی ارشد	دکترا
تعداد	۱۵۲۳۹	۱۶۸۴۳	۱۲۵۳	۱۷۹۱	۱۷۲	۳۸

سالنامهٔ آمار سال ۱۳۷۶



تمرین

۱ــ نمودار دایره‌ای برای چه نوع متغیری مناسب است؟

۲ــ اگر فراوانی داده‌ها را دوبرابر کنیم آیا زاویه مرکزی عوض می‌شود؟ توضیح دهید.

۳ــ در یک نمودار دایره‌ای اگر یک زاویهٔ مرکزی دوبرابر زاویه مرکزی دیگر باشد چه رابطه‌ای بین فراوانی‌های این دو زاویه وجود دارد؟

۴ــ آیا ترتیب کنار هم قرارگرفتن نواحی در نمودار دایره‌ای اهمیت دارد؟ توضیح دهید.

۵ ــ گروه خونی همکلاسی‌های خود را بپرسید و برای اطلاعات جمع‌آوری شده نمودار دایره‌ای رسم کنید.

۶ــ با توجه به جدول زیر، نمودار دایره‌ای توزیع خانوارها برحسب جمعیت را رسم کنید.

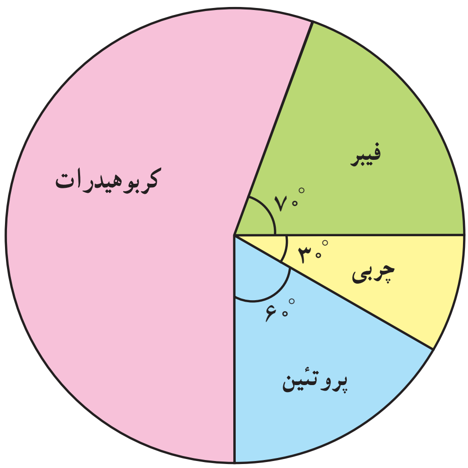
نوع خانوار	کم‌جمعیت	با جمعیت متوسط	پرجمعیت	جمع
تعداد خانوار	۱۱	۲۳	۶	۴۰

۷ــ اگر شعاع دایره را تغییر دهیم آیا تفسیر جدیدی از توزیع داده‌ها به‌دست می‌آید یا خیر؟

۸ــ نمودار دایره‌ای زیر سهم وزنی ترکیبات تشکیل‌دهندهٔ یک بسته غذای کنسرو شده را نشان می‌دهد.

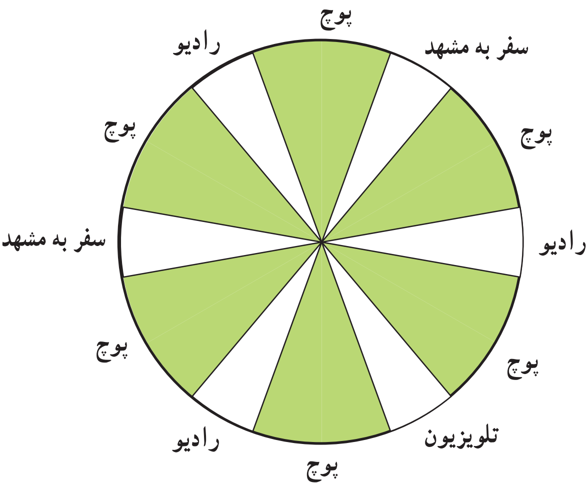
الف ــ چه کسری از این ترکیبات :

چربی است؟ کربوهیدرات است؟



ب ــ چند گرم پروتئین در یک بسته ۳۶ گرمی از این محصول وجود دارد؟

۹ــ در یک مسابقه علمی، شرکت‌کنندگان پس از دادن پاسخ درست به سؤال با گرداندن یک صفحه‌گردان، جایزه دریافت می‌کنند. اگر نواحی رنگ شده °۴۰ و نواحی بدون رنگ °۲۰ باشند به سؤالات زیر پاسخ دهید :



ــ امکان بردن جایزه بیش‌تر است یا نبردن جایزه؟

ــ چند درصد شانس بردن رادیو وجود دارد؟

ــ چند درصد شانس بردن تلویزیون وجود دارد؟

ــ چند درصد شانس بردن سفر به مشهد وجود دارد؟

ــ چند درصد امکان نبردن جایزه وجود دارد؟

ــ نمودار دایره‌ای نشان‌دهندهٔ بردن جایزه را رسم کنید.

۱۰ــ مخارج علی در سال اول دانشگاه در جدول روبرو آمده است :

نمودار دایره‌ای هزینه‌های علی را رسم کنید.

مخارج	درصد
غذا	٪۳۰
اجاره	٪۲۵
لباس	٪۱۵
کتاب	٪۱۰
دیگر مخارج	٪۲۰

۱۱ــ جدول زیر اطلاعات راجع به سطح زیرکشت و میزان تولید مرکبات کشور در سال ۷۳ را برای استان‌های ذکرشده نشان می‌دهد.

استان	سطح زیر کشت (هکتار) نهال	بارور	میزان تولید (تن)	عمل‌کرد (کیلوگرم)
مازندران	۵۶۲۰	۷۶۷۷۵	۱۳۷۳۶۸۰	۱۷۸۹۲
فارس	۳۲۱۸	۳۹۱۵۱	۶۱۷۶۳۰	۱۵۷۷۶
کرمان	۱۶۹۱	۱۱۰۴۹	۹۴۸۷۶	۸۵۷۸
هرمزگان	۵۴۷۹	۲۴۲۶۰	۳۴۱۲۳۹	۱۳۸۴۹

آمارنامه کشاورزی، سال ۷۳ ــ اداره کل آمار و اطلاعات، معاونت طرح و برنامه وزارت کشاورزی، نشریهٔ شمارهٔ ۱۸، دی‌ماه ۷۴.

نمودار دایره‌ای مربوط به میزان تولید مرکبات بر حسب تن را برای این استانها رسم کنید.



نمودار ساقه و برگ

آیا تاکنون فکر کرده‌اید که می‌توان از اعداد در تهیه یک نمودار استفاده کرد؟

آیا نموداری دیده‌اید که اعداد، شکل‌دهنده آن باشند؟

در یک بررسی که روی ۱۵۰۰ نفر انجام شده‌بود اعلام شد که ۴۵٪ صبحها قبل از ساعت ۸ صبح از خانه بیرون می‌روند. برای بررسی درستی این خبر، تحقیقی روی ۲۴ گروه هر یک شامل ۱۵۰۰ نفر انجام شد. درصد افرادی که در هر گروه قبل از ساعت ۸ صبح از خانه بیرون می‌رفتند به شکل زیر گزارش داده شد :

۴۶	۴۵	۴۶	۴۵	۴۵	۴۴
۴۴	۴۶	۴۴	۴۲	۴۷	۴۳
۴۵	۴۳	۴۸	۴۴	۴۷	۴۰
۴۷	۴۸	۴۳	۴۶	۴۵	۴۲

همان‌طوری‌که می‌بینید تمام این اعداد در رقم ۴ مشترک هستند. یعنی تماماً اعداد طبیعی در فاصلهٔ ۴۰ تا ۵۰ هستند، از این‌رو بخش اصلی این اعداد را که همان ۴۰ است می‌توانیم کنار بگذاریم و سایر ارقام را به‌صورت زیر بنویسیم :

۴۰+ ۰, ۲, ۲, ۳, ۳, ۳, ۴, ۴, ۴, ۴, ۵, ۵, ۵, ۵, ۵,
۶, ۶, ۶, ۶, ۷, ۷, ۷, ۸, ۸

البته داده‌ها را در بالا از کوچک به بزرگ مرتب کرده‌ایم. یک راه ساده‌تر نوشتن اعداد فوق آن است که مثلاً عدد ۴۷ را به‌صورت ۴ ۷ بنویسیم (کمی فاصله بین ۴ و ۷). در این صورت عدد چهار را ثابت نگاه داشته و سایر ارقام را در مقابل آن به‌ترتیب می‌نویسیم. در مثال مورد بحث چون تعداد زیادی داده داریم که در عدد ۴ مشترک هستند می‌توانیم هر رقم را در یک ردیف به‌صورت روبرو بنویسیم.

۴ ۰
۲ ۲
۳ ۳ ۳
۴ ۴ ۴ ۴
۵ ۵ ۵ ۵ ۵
۶ ۶ ۶ ۶
۷ ۷ ۷
۸ ۸

شکل ۱۸

ملاحظه می‌کنید که این‌طور مرتب کردن اعداد، یک نمودار میله‌ای از داده‌ها مشخص می‌کند (برای آن‌که بهتر این نکته را ببینید می‌توانید صفحه کتاب را ۹۰ درجه دوران دهید). خوبی این نمودار در این است که تمام داده‌ها را دربر دارد. عدد چهار را که در واقع همان ۴۰ است و قسمت اصلی داده‌هاست **ساقه** می‌گوییم. سایر ارقام را که متصل به این ساقه هستند **برگ** می‌گوییم و شکلی که در صفحهٔ قبل حاصل شد **نمودار ساقه و برگ** می‌گویند. در زیر شما چند نمونه از نمودارهای ساقه و برگ را ملاحظه می‌کنید.

مثال: بعضی از مردم نسبت به گزیدگی پشه حساس‌اند. محل گزیدگی در این افراد تا مدتی همراه با خارش و سوزش است که پس از مدتی این آثار برطرف می‌شود. در این‌جا بخشهای مسأله آماری، عبارتند از :

جامعه: تمام کسانی که نسبت به گزیدگی پشه حساس‌اند.

متغیر تصادفی: مدت زمانی که طول می‌کشد تا آثار حساسیت محو شود.

نمونه: برای مطالعهٔ این مدت زمان چون تمام افراد در اختیار ما نیستند لذا تعدادی از این افراد مثلاً ۴۰ نفر آنها را مورد مطالعه قرار داده‌ایم. داده‌های حاصل از این نمونه عبارتند از : (داده‌ها برحسب دقیقه اندازه‌گیری شده‌اند).

۱۰/۱	۱۶/۵	۱۲/۷	۱۱/۴	۱۵/۰	۹/۹	۱۱/۲	۱۰/۵
۸/۱	۱۰/۹	۸/۳	۷/۹	۶/۲	۱۱/۶	۱۱/۴	۱۲/۷
۱۰/۴	۹/۱	۱۱/۲	۱۲/۵	۸/۴	۱۱/۷	۱۰/۵	۳/۸
۸/۶	۷/۴	۸/۸	۱۱/۴	۵/۹	۱۲/۳	۱۳/۴	۹/۱
۹/۹	۱۲/۹	۹/۸	۱۰/۹	۱۱/۵	۱۱/۵	۱۴/۷	۱۳/۶



۰۳
۰۴
۰۵
۰۶
۰۷
۰۸
۰۹
۱۰
۱۱
۱۲
۱۳
۱۴
۱۵
۱۶

برای یکنواختی، داده‌ها را به‌صورت یک عدد سه رقمی می‌نویسیم، مثلاً عدد ۳/۸ را به‌صورت ۳۸۰ و عدد ۱۶/۵ را به‌صورت ۱۶۵ می‌نویسیم، اگر عددی قسمت اعشاری نداشت مثلاً به‌صورت ۱۲ بود آن را به‌صورت ۱۲۰ می‌نویسیم. با این مقدمه نمودار ساقه و برگ را به‌صورت روبرو تشکیل می‌دهیم. ابتدا ساقه را تشکیل می‌دهیم، ساقه را به‌صورت دو رقم اول داده‌ها در نظر گرفته‌ایم. بنابراین ساقه به‌صورت روبرو است :

حال برای تشکیل برگها، هر داده را در محل مربوطه روی ساقه قرار می‌دهیم، مثلاً مقدار ۱۱۲ (مربوط به داده ۱۱/۲) را در محل ۱۱ روی ساقه قرار می‌دهیم، و آن را به‌صورت ۲ ۱۱ می‌نویسیم ۱۱/۵ در همان محل و به‌صورت ۵ ۱۱ نوشته می‌شود. پس این دو دادهٔ اخیر در محل ۱۱ به‌صورت زیر نوشته می‌شوند :

۱۱ ۲۵

با ادامهٔ این عمل برگ مربوط به اتصال ۱۱ روی ساقه به‌صورت زیر تکمیل می‌شود :

۱۱ ۲۲۴۴۴۵۵۶۷

این سطر نشان می‌دهد که ۹ داده داریم که با ۱۱ شروع می‌شوند (داده‌های ۱۱/۲،۱۱/۲، ۱۱/۴، ۱۱/۴، ۱۱/۴، ۱۱/۵، ۱۱/۵، ۱۱/۶، ۱۱/۷) با تکرار این عمل برای سایر مقادیر، نمودار ساقه و برگ به‌صورت زیر تکمیل می‌شود :

۰۳	۸
۰۴	
۰۵	۹
۰۶	۲
۰۷	۴۹
۰۸	۱۳۴۶۸
۰۹	۱۱۸۹۹
۱۰	۱۴۵۵۹۹
۱۱	۲۲۴۴۴۵۵۶۷
۱۲	۳۵۷۷۹
۱۳	۴۶
۱۴	۷
۱۵	۰
۱۶	۵

کلید نمودار : ۷/۴ = ۴ ۷

در یک تست انگلیسی ۸۰ سؤاله نتایج زیر توسط دانش‌آموزان یک کلاس بدست آمده است. یک نمودار ساقه و برگ برای داده‌های زیر رسم می‌کنیم :

۲۵	۳۰	۳۰	۳۷	۳۸	۳۸	۴۴	۴۴	۴۴
۴۹	۵۰	۵۰	۵۶	۵۷	۶۱	۳۱	۳۱	۳۲
۳۲	۳۲	۳۵	۳۵	۳۷	۳۹	۴۰	۴۰	۴۰
۴۰	۴۱	۴۲	۴۵	۴۵	۴۶	۴۶	۴۶	۴۷
۴۷	۴۷	۴۹	۵۰	۵۰	۵۰	۵۱	۵۲	۵۲
۵۲	۵۴	۵۶	۵۶	۵۶	۵۷	۶۱	۶۴	۶۶
۶۷	۷۱	۷۸						

در اینجا برای تشکیل ساقه‌ها از رقم دهگان و برای تشکیل برگها از رقم یکان استفاده می‌کنیم، مثلاً عدد ۷۸ را به‌صورت

۷ ۸

می‌نویسیم که ۷ روی ساقه و ۸ روی برگها قرار دارد. با این ترتیب نمودار زیر به‌دست می‌آید :

ساقه	برگ
۲	۵
۳	۰ ۰ ۱ ۱ ۲ ۲ ۲ ۵ ۵ ۷ ۷ ۸ ۸ ۹
۴	۰ ۰ ۰ ۰ ۱ ۲ ۴ ۴ ۴ ۵ ۵ ۶ ۶ ۶ ۷ ۷ ۷ ۹ ۹
۵	۰ ۰ ۰ ۰ ۰ ۱ ۲ ۲ ۲ ۴ ۶ ۶ ۶ ۶ ۷ ۷
۶	۱ ۱ ۴ ۶ ۷
۷	۱ ۸

کلید نمودار : ۵۲ = ۲ ۵

برای انجام یک پروژه در کلاس دو هفته در نظر گرفته شده است. تعداد روزهایی که هر دانش‌آموز صرف انجام پروژه کرده است در زیر آمده است :

۷	۷	۸	۸	۸	۸	۹	۹	۹	۹	۹
۹	۹	۹	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰
۱۰	۱۰	۱۰	۱۰	۱۰	۱۰	۱۱	۱۱	۱۱	۱۱	۱۱
۱۱	۱۱	۱۱	۱۱	۱۲	۱۲	۱۲	۱۲	۱۲	۱۳	۱۳
۱۴										

نمودار ساقه و برگ به این شکل خواهد بود :

۰	۷ ۷ ۸ ۸ ۸ ۸
۰	۹ ۹ ۹ ۹ ۹ ۹ ۹ ۹
۱	۰ ۰ ۰ ۰ ۰ ۰ ۰ ۰ ۰ ۰ ۰ ۰ ۰ ۰
۱	۱ ۱ ۱ ۱ ۱ ۱ ۱ ۱ ۱

کلید نمودار : ۱۰ = ۰ ۱

فکر می‌کنید چرا برگها در این نمودار در چند سطر آورده شده است؟

تمرین

۱ــ نمودار ساقه و برگ روبرو داده شده است :

ــ داده‌های موجود در این نمودار را بنویسید.

ساقه	برگ
۱	۰ ۳ ۳ ۴
۲	۰ ۲ ۴ ۸ ۸
۳	۲

۲ــ در نمودار ساقه و برگ روبرو چه اعدادی می‌توانند در □ قرار گیرند ؟

ساقه	برگ
۱	۲□ ۵ ۶
۲	۷ ۹
۳	۲ ۴□

۳ــ نمودار ساقه و برگ ارتفاع ۲۱۸ آتش‌فشان به‌شکل زیر است.

۰	۲ ۵ ۶ ۶ ۶ ۷ ۸ ۹
۱	۰ ۱ ۳ ۶ ۷ ۷ ۹ ۹
۲	۰ ۰ ۰ ۱ ۱ ۲ ۲ ۲ ۴ ۴ ۴ ۵ ۵ ۶ ۶ ۶ ۷ ۷ ۸ ۸ ۹ ۹ ۹
۳	۰ ۱ ۱ ۲ ۲ ۴ ۴ ۵ ۵ ۵ ۵ ۶ ۶ ۶ ۷ ۸ ۹ ۹
۴	۰ ۱ ۱ ۱ ۲ ۳ ۳ ۳ ۳ ۳ ۴ ۴ ۴ ۶ ۷ ۸ ۸ ۹ ۹ ۹ ۹ ۹
۵	۰ ۰ ۱ ۱ ۲ ۲ ۲ ۳ ۴ ۴ ۵ ۵ ۶ ۶ ۶ ۶ ۶ ۶ ۷ ۷ ۷ ۹ ۹
۶	۰ ۰ ۱ ۱ ۴ ۴ ۵ ۵ ۶ ۶ ۶ ۶ ۷ ۷ ۷ ۸ ۸ ۹
۷	۰ ۰ ۰ ۰ ۱ ۱ ۱ ۲ ۳ ۳ ۴ ۵ ۵ ۵ ۶ ۷ ۸ ۸ ۹
۸	۱ ۲ ۲ ۲ ۲ ۳ ۳ ۳ ۵ ۶ ۷ ۹
۹	۰ ۰ ۰ ۱ ۲ ۳ ۳ ۴ ۴ ۵ ۵ ۶ ۷ ۷ ۹
۱۰	۰ ۱ ۱ ۲ ۲ ۳ ۳ ۴ ۴ ۵ ۶ ۸ ۹
۱۱	۰ ۱ ۱ ۲ ۳ ۳ ۴ ۶ ۶ ۹
۱۲	۱ ۱ ۲ ۴ ۴ ۴ ۵ ۶
۱۳	۰ ۳ ۴ ۷ ۸
۱۴	۰ ۰
۱۵	۶ ۶ ۷
۱۶	۲ ۵
۱۷	۲ ۹
۱۸	۵
۱۹	۰ ۳ ۳ ۷ ۹

(سانتی‌متر ۳۰ = ۱ فوت)

واحد : ۱۰۰ فوت

فوت ۹۰۰ = ۹ ۰

فوت ۱۹۳۰۰ = ۳ ۱۹

ــ ارتفاع بلندترین آتش‌فشان چه‌قدر است؟

ــ چند آتش‌فشان ارتفاعی بین ۱۰۰۰۰ و ۱۲۰۰۰ فوت دارند؟

ــ چند آتش‌فشان با ارتفاع ۴۳۰۰ فوت وجود دارد؟

۴ــ داده‌های روبرو، زمان صرف شده برحسب دقیقه برای رسیدن به محل کار ۲۰ کارمند را نشان می‌دهد.

ــ نمودار ساقه و برگ این داده‌ها را رسم کنید.

۱۰	۵۰	۶۵	۳۳
۴۸	۵	۱۱	۲۳
۳۷	۲۶	۲۶	۳۲
۱۷	۷	۱۳	۱۹
۲۹	۴۳	۲۱	۲۲

۵ــ نمرات تست زبان دو کلاس (از ۱۰۰) در نمودار ساقه و برگ در زیـر نشان داده شده است :

ــ کم‌ترین نمره در هر یک از کلاس‌های دختران و پسران چند است؟

ــ چند نفر در کلاس دختران و چند نفر در کلاس پسران نمرهٔ بالاتر از ۶۰ کسب کرده‌اند؟

ــ جدول فراوانی داده‌های فوق را با توجه به دسته‌بندی روبرو رسم کنید.

F ← ۰ ≤ X < ۶۰

D ← ۶۰ ≤ X < ۷۰

C ← ۷۰ ≤ X < ۸۰

B ← ۸۰ ≤ X < ۹۰

A ← ۹۰ ≤ X ≤ ۱۰۰

برگ کلاس پسران	ساقه	برگ کلاس دختران
	۲	۹
۸ ۰	۳	۴
۹ ۴	۴	۷ ۸
۹ ۱ ۰ ۰	۵	۰ ۱ ۸
۹ ۸ ۸ ۸ ۶ ۰ ۰	۶	۰ ۲ ۲ ۴ ۸ ۸ ۹
۹ ۸ ۸	۷	
۶ ۵ ۵ ۵	۸	۱ ۲ ۲ ۴ ۵ ۵ ۸ ۹
۸ ۷ ۷ ۴ ۴ ۱ ۰	۹	۳ ۴ ۴ ۴ ۷ ۷
۰	۱۰	۰ ۰

ــ اگر شاخص‌های D, C, B, A و F برای نتایج امتحان استفاده شود، نمودار میله‌ای را برای هر یک از کلاس‌ها رسم کنید.

ــ نمودار مستطیلی هر کلاس را رسم کنید.

ــ کدام کلاس درصد قبولی بالاتری دارد؟ (نمرات C, B, A و D قبول می‌باشند)

۶ ــ مربی تیم بسکتبال یک مدرسه جهت تشکیل تیم بسکتبال مدرسه، طول قد دانش‌آموزان دو کلاس را اندازه‌گیری نموده است. طول قدهای دانش‌آموزان این دو کلاس در جدول‌های زیر آمده است :

کلاس الف	کلاس ب
۱۳۴, ۱۱۹, ۱۳۷, ۱۴۰, ۱۶۳,	۱۴۴, ۱۴۲, ۱۲۶, ۱۴۳, ۱۵۴,
۱۴۷, ۱۳۵, ۱۶۸, ۱۳۱, ۱۶۰,	۱۶۴, ۱۶۵, ۱۶۶, ۱۲۸, ۱۵۷,
۱۱۶, ۱۴۴, ۱۴۶, ۱۳۳, ۱۵۱,	۱۵۶, ۱۲۹, ۱۴۰, ۱۲۸, ۱۴۷,
۱۳۰, ۱۳۷, ۱۳۷, ۱۲۵, ۱۲۹,	۱۵۹, ۱۲۲, ۱۵۵, ۱۴۶, ۱۶۹,
۱۴۲, ۱۶۹, ۱۴۷, ۱۱۷,	۱۱۲, ۱۵۸, ۱۳۱, ۱۳۳, ۱۴۴

الف ــ نمودار ساقه و برگ برای هر یک از جدول‌ها را رسم کنید.

ب ــ کوتاه‌قدترین دانش‌آموز در کدام کلاس است؟



ج ــ بلندقدترین دانش‌آموز در کدام کلاس است؟

د ــ قد چند دانش‌آموز در کلاس الف کوتاه‌تر از ۱۱۰ سانتی‌متر است؟

هـ ــ اگر شرط عضویت در تیم بسکتبال مدرسه قد حداقل ۱۲۰ سانتی‌متر باشد، چه تعداد از دانش‌آموزان هر کلاس واجد شرط عضویت در تیم بسکتبال می‌باشند؟

پروژه

به‌عنوان پروژهٔ آماری، محمد می‌بایست یک نمونهٔ ۳۰ تایی ازماشین‌هایی که در یک پارکینگ در کنار یک ورزشگاه بزرگ پارک شده بودند انتخاب نموده و نام ماشین و سال ماشین را به‌دست آورد.

الف ــ اگر در این پارکینگ ۹۵۰ ماشین پارک شده باشد، توضیح دهید که محمد با استفاده از اعداد تصادفی چگونه می‌تواند نمونه خود را انتخاب کند.

ب ــ پس از این‌که محمد نمونه خود را انتخاب کرد، او متوجه شد که ۸ ماشین پیکان، ۷ ماشین پژو، ۵ ماشین رنو، ۳ ماشین تویوتا، ۳ ماشین فولکس، ۳ ماشین دوو و یک ماشین بنز بود. جدول فراوانی داده‌های به‌دست‌آمده را تشکیل دهید و نمودار میله‌ای آن را رسم کنید.

ــ نمودار دایره‌ای این نمونه را بر اساس درصد فراوانی ماشین‌ها رسم کنید.

ــ در قسمت ب، چه نوع متغیری داریم؟ آیا می‌توانیم برای این متغیر نمودار مستطیلی رسم کنیم؟

ــ سال ماشین‌ها چه نوع متغیری است؟ آیا می‌توان برای این متغیر نمودار مستطیلی رسم کرد؟

رسم نمودارها با استفاده از MINITAB



در این بخش چگونگی رسم نمودار میله‌ای، مستطیلی، دایره‌ای و ساقه و برگ را خواهید آموخت.

نمودار دایره‌ای و نمودار میله‌ای

مثال ۱: جدول فراوانی رشته‌های تحصیلی ۲۵ دانشجو در دست است.
با استفاده از MINITAB نمودار میله‌ای و دایره‌ای این اطلاعات را رسم کنید.

حل: اگر از MINITAB WINDOWS استفاده می‌کنید مراحل زیر را دنبال کنید :

تعداد دانشجویان	رشته تحصیلی
۶	اقتصاد
۳	ریاضی
۶	فیزیک
۲	شیمی
۸	ادبیات

قدم اول: رشته‌های تحصیلی و فراوانی دانشجویان هر رشته را در ستون‌های C۱ و C۲ وارد کنید.

قدم دوم: روی GRAPH در لیست انتخاب کلیک کنید.

قدم سوم: روی CHART کلیک کنید.

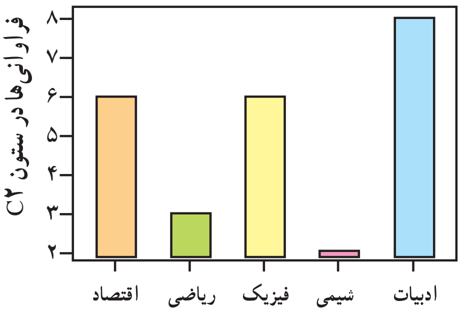
قدم چهارم: در کادر، زیر Y، تایپ کنید C۲ و زیر X، تایپ کنید C۱

قدم پنجم: روی OK کلیک کنید.

اگر از فرمان‌های زبانی MINITAB استفاده می‌کنید، ابتدا رشته‌های تحصیلی و فراوانی دانشجویان را در ستون‌های C۱ و C۲ با استفاده از فرمان SET وارد کنید. توجه داشته باشید برای واردکردن رشته‌ها باید از فرمان FORMAT استفاده کنید. سپس فرمان زیر را تایپ کنید :

MTB > CHART C2 * C1

بدون در نظرگرفتن این که از کدام طریق شما کار را به پایان رسانیده‌اید، نمودار میله‌ای زیر را خواهید داشت.



برای رسم نمودار دایره‌ای چنین عمل کنید :

قدم اول: رشته‌های تحصیلی و فراوانی دانشجویان هر رشته را در ستون‌های C۱ و C۲ وارد کنید.

قدم دوم: روی GRAPH در لیست انتخاب کلیک کنید.

قدم سوم: روی PIE CHART کلیک کنید.

قدم چهارم: روی CHART TABLE کلیک کنید. سپس کنار CATEGORIES IN در کادر، C۱ تایپ کنید و در کنار FREQUENCIES IN در کادر C۲ تایپ کنید.

قدم پنجم: روی OK کلیک کنید.

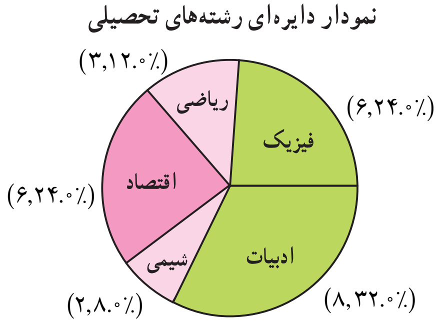
اگر از فرمان‌های زبانی MINITAB استفاده می‌کنید ابتدا رشته‌های تحصیلی و فراوانی دانشجویان را در ستون‌های C۱ و C۲ وارد کنید. سپس فرمان‌های زیر را تایپ کنید.

MTB > ٪ PIE C۱ ;

SUBC > COUNTS C2 .

اولین فرمان از MINITAB می‌خواهد نمودار دایره‌ای برای رشته‌های موجود در ستون C۱ رسم کند و زیر فرمان COUNTS فراوانی هر یک از رشته‌ها را که در ستون C۲ آمده است ارائه می‌دهد.

بدون در نظرگرفتن این که شما از چه روشی کار را به پایان رسانیده‌اید، نمودار دایره‌ای زیر رسم می‌شود.



نمودار مستطیلی

مثال ۲: در جدول زیر قد ۳۰ بازیکن بسکتبال به‌طور تصادفی انتخاب شده‌اند (برحسب اینچ، ۱ اینچ = ۲/۵ سانتی‌متر است) آمده است.

۸۲	۷۷	۷۴	۷۳	۷۶	۷۹	۸۴	۸۱
۷۲	۷۸	۸۲	۸۰	۷۶	۷۶	۸۱	۷۵
۸۴	۷۹	۷۸	۷۸	۷۷	۸۰	۸۳	۸۰
		۷۹	۸۳	۷۹	۸۳	۸۶	۷۳

جدول فراوانی داده‌ها با فرض داشتن ۵ دسته به‌شکل زیر خواهد بود.

مرکز داده‌ها	f	قد برحسب اینچ
۷۳/۵	۴	۷۲ـ۷۵
۷۶/۵	۶	۷۵ـ۷۸
۷۹/۵	۱۰	۷۸ـ۸۱
۸۲/۵	۷	۸۱ـ۸۴
۸۵/۵	۳	۸۴ـ۸۷

طول دسته‌ها = ۳

برای رسم نمودار مستطیلی داده‌های فوق با استفاده از MINITAB چنین عمل کنید :

قدم اول: داده‌ها را در ستون C۱ وارد کنید.

قدم دوم: روی GRAPH در لیست انتخاب، کلیک کنید.

قدم سوم: روی HISTOGRAM کلیک کنید.

قدم چهارم: C۱ را در کادر، زیر X تایپ کنید.

قدم پنجم: روی OPTION در پایین این پنجره کلیک کنید.

قدم ششم: روی دایره‌ای که در کنار MID POINT زیر ________ TYPE OF INTERVALS قرار دارد کلیک کنید. سپس روی دایره که در کنار MIDPOINT/CUT POINT POSITION زیر ________ DEFINITION OF INTERVALS قرار دارد کلیک کنید. سپس در کادری که در کنار MIDPOINT / CUT POINT POSITION قرار دارد تایپ کنید:

73.5 : 85.5/3 . توجه داشته باشید که ۷۳/۵ مرکز دستهٔ اول و ۸۵/۵ مرکز دسته آخر و ۳ طول دسته‌ها می‌باشد.

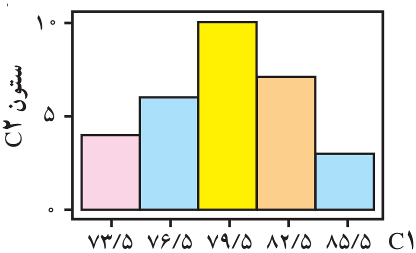
قدم هفتم: روی OK در پایین صفحه کلیک کنید. سپس مجدداً روی OK کلیک کنید.

نمودار مستطیلی ظاهر می‌شود.

اگر از فرمان‌های زبانی MINITAB استفاده می‌کنید، مرکز دسته‌ها را در ستون C۱ با استفاده از فرمان SET وارد کنید، سپس فرمان‌های زیر را تایپ کنید:

```
MTB > HISTOGRAM  C1  ;
SUBC > MIDPOINTS   73.5:85.5/3.
```

توجه داشته باشید از "؛" و "." استفادهٔ صحیح همان‌گونه که قبلاً توضیح داده شد، استفاده شود. بدون توجه به روش استفاده شده، نمودار مستطیلی زیررسم خواهد شد:



در مثال بالا دسته‌ها را خودمان انتخاب کردیم. اگر می‌خواهید MINITAB خودش دسته‌ها و طول دسته‌ها را انتخاب کند در MINITAB WINDOWS قدم پنجم و ششم را اجرا نکنید و در فرمان‌های زبانی MINITAB فقط HISTOGRAM C۱ تایپ کنید و زیر فرمان را تایپ نکنید.

اگر می‌خواهید نمودار چندبر فراوانی با استفاده از MINITAB WINDOWS رسم کنید، روی GRAPH در لیست انتخاب کلیک کنید و سپس HISTOGRAM را انتخاب کنید و زیر X در کادر، C۱ تایپ کنید. روی DISPLAY کلیک کنید. روی CONNECT کلیک کنید، روی OPTION در پایین کادر کلیک کنید، روی MIDPOINT/CUT POINT POSITIONS کلیک کنید، در کادر73.5:85.5/3 تایپ کنید و در پایان روی OK دوبارکلیک کنید. MINITAB چندبر فراوانی را رسم خواهد کرد.

اگر ازفرمان‌های زبانی MINITAB استفاده می‌کنید، فرمان‌های زیر را تایپ کنید :

```
MTB > HISTOGRAM  C1 ;
SUBC > MIDPOINTS  73.5:85.5/3 ;
SUBC > CONNECT.
```

نمایش نمودار ساقه و برگ

نمرات زبان ۳۰ دانش‌آموز در یک مؤسسه زبان (از ۱۰۰ نمره) به‌قرار زیر است :

۷۵ , ۵۲ , ۸۰ , ۹۶ , ۶۵ , ۷۹ , ۷۱ , ۸۷ , ۹۳ , ۹۵
۶۹ , ۷۲ , ۸۱ , ۶۱ , ۷۶ , ۸۶ , ۷۹ , ۶۸ , ۵۰ , ۹۲
۸۳ , ۸۴ , ۷۷ , ۶۴ , ۷۱ , ۸۷ , ۷۲ , ۹۲ , ۵۷ , ۹۸

برای نمایش نمودار ساقه و برگ این ۳۰ دانش‌آموز با استفاده از MINITAB WINDOWS مراحل زیر را انجام دهید :

قدم اول: داده‌ها را در ستون C۱ وارد کنید.

قدم دوم: روی GRAPH در لیست انتخاب کلیک کنید.

قدم سوم: روی CHARACTER GRAPHS کلیک کنید.

قدم چهارم: روی STEM - AND - LEAF کلیک کنید.

قدم پنجم: C۱ را در کادر زیر Variable و ۱۰ را در کادر در کنار Increment تایپ کنید.

قدم ششم: روی OK کلیک کنید. نمودار ساقه و برگ نمایش داده خواهد شد.

اگر شما از فرمان‌های زبانی MINITAB استفاده می‌کنید ابتدا داده‌ها را با استفاده از فرمان SET در C۱ وارد کنید و سپس فرمان‌های زیر را تایپ کنید :

```
MTB > STEM – AND – LEAF   C۱;
SUBC > INCREMENT = 10.
```

در زیر فرمان INCREMENT ، ۱۰ فاصله بین دو ساقهٔ متوالی در نمودار را نشان می‌دهد.

بدون درنظرگرفتن روش کار، نمودار ساقه و برگ زیر نمایش داده می‌شود :

```
STEM – AND – LEAF  OF  C1   N = 30
LEAF    UNIT  =  1.0
                         3     5   0 2 7
                         8     6   1 4 5 8 9
                        (9)    7   1 1 2 2 5 6 7 9 9
                        13     8   0 1 3 4 6 7 7
                         6     9   2 2 3 5 6 8
```

در این نمایش N = ۳۰ تعداد داده‌ها را نشان می‌دهد. LEAF UNIT = 1.0 نشان می‌دهد که ` . ´ ، علامت اعشاری، پس از اولین رقم بعد از برگ قرار می‌گیرد به‌طور مثال اولین عدد در این نمودار ۵۰.۰ و دومین عدد ۵۲.۰ ... خواهد بود.

اگر Leaf Unit =.01 باشد، اولین عدد 5.0 و دومین عدد 5.2 و سومین عدد 5.7 ... خواهد بود.

از طرف دیگر اگر LEAF UNIT =۱۰ باشد، اولین عدد ۵۰۰، دومین عدد ۵۲۰ و سومین عدد ۵۷۰ ... خواهند بود. عددی که داخل پرانتز در ستون اول، MINITAB چاپ می‌کند (در این‌جا (۹)) نشان‌دهنده ردیفی است که میانه در آن قرار دارد. عدد ۹ نشان‌دهنده تعداد برگ‌های ساقه در این ردیف می‌باشد. اعداد بالا و پایین این عدد فراوانی تجمعی دسته‌های قبل و بعد از این دسته را نشان می‌دهند.

تمرین

۱ــ فرض کنید از ۵۰ دانش‌آموز دبیرستانی پرسیده شده است که سال چندم هستند. جدول زیر نتیجه این بررسی را نشان می‌دهد.

اول	سوم	اول	اول	سوم	پیش‌دانشگاهی	دوم	اول
دوم	دوم	دوم	اول	دوم	دوم	اول	اول
پیش‌دانشگاهی	سوم	دوم	اول	اول	سوم	پیش‌دانشگاهی	سوم
سوم	سوم	پیش‌دانشگاهی	سوم	اول	دوم	دوم	پیش‌دانشگاهی
سوم	دوم	پیش‌دانشگاهی	اول	اول	پیش‌دانشگاهی	پیش‌دانشگاهی	اول
اول	دوم	پیش‌دانشگاهی	اول	سوم	دوم	سوم	پیش‌دانشگاهی
						پیش‌دانشگاهی	پیش‌دانشگاهی

ــ جدول فراوانی این داده‌ها را تشکیل دهید.

ــ نمودار میله‌ای و دایره‌ای این داده‌ها را رسم کنید. (با استفاده از MINITAB)

۲ــ داده‌های زیر وزن ۲۰ بستهٔ پستی را برحسب کیلوگرم که در هفته گذشته از یکی از شعبه‌های اداره پست، ارسال شده‌اند نشان می‌دهد.

۲/۵	۱/۹	۹/۳	۵/۱	۳/۴	۸/۲	۷/۵	۱/۸
۰/۷	۸/۵	۶/۳	۲/۰	۸/۶	۶/۲	۵/۸	۷/۳
				۳/۷	۵/۲	۷/۳	۳/۸

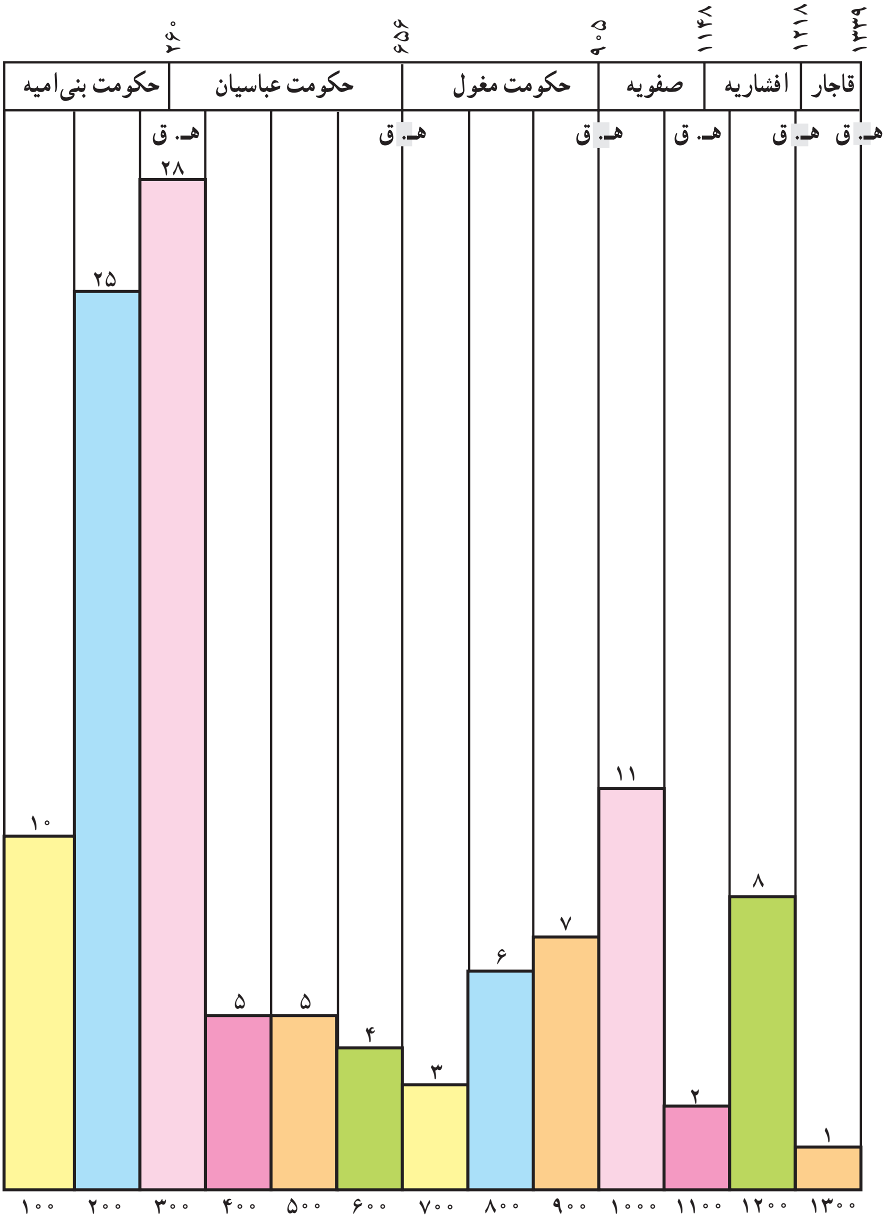
ــ با استفاده از MINITAB ، نمودار ساقه و برگ این داده‌ها را با درنظرگرفتن INCREMENT =۱.۰ رسم کنید. مقادیر برگ‌ها را مشاهده کنید.

۳ــ داده‌های زیر تعداد صفحه کلیدهای جمع‌شده را در یک شرکت الکترونیکی نشان می‌دهد.

۴۸	۴۲	۴۴	۴۶	۵۶	۴۱	۴۸	۵۲	۴۵
۵۰	۵۲	۴۳	۴۶	۴۸	۵۱	۵۳	۵۱	۵۳
		۵۲	۴۹	۵۰	۴۷	۴۴	۴۷	۵۴

ــ جدول فراوانی این داده‌ها را با داشتن ۴ دسته تشکیل دهید.

ــ نمودار مستطیلی و چندبر فراوانی این داده‌ها را با استفاده از MINITAB رسم کنید.



نمودار تعداد دانشمندان ریاضی و
نجوم ایران در هر صد سال از قرن اوّل
هجری تا پایان قرن ۱۳ هـ. ق

نقل از: نگارنامه تاریخی و
زندگی‌نامه دانشمندان ریاضی و
نجوم ایران در دوره اسلام

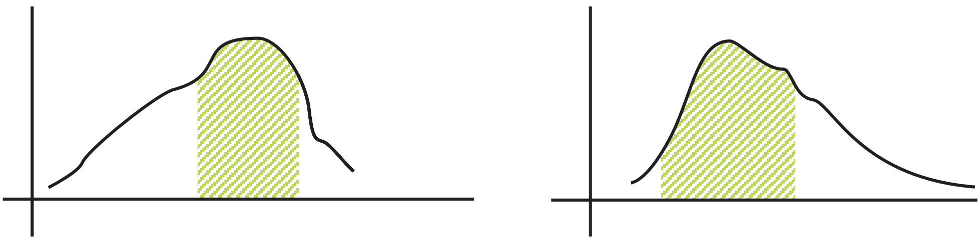


آفرینش را آغاز کرد و آفریدگان را به یکبار پدید آورد، بی‌آنکه اندیشه‌ای به‌کار برد، یا از آزمایشی سود برد یا جنبشی پدید آرد یا کسی را به خدمت گمارد. از هر چیز بهنگام بپرداخت، و اجزای مخالف را با هم سازوار ساخت، و هر طبیعت را اثری بداد و آن اثر را در ذات آن نهاد. از آن پیش که بیافریند به آفریدگان دانابود. بر آغاز و انجامشان بینا، و با سرشت و چگونگی آنان آشنا.

فصل ششم

شاخص‌های مرکزی

در فصل‌های قبل اشاره کردیم که یک اظهارنظر را وقتی می‌توان علمی تلقی کرد که همراه با عدد و رقم باشد. تا به حال ما به‌جز دامنهٔ تغییرات، کمیّتی که بتواند جامعه را معرفی کند ارائه نکرده‌ایم، آن‌چه که تا به حال مطرح شده است، دربارهٔ کم کردن گوناگونی (از طریق دسته‌بندی آن‌ها) و یا نمایش هندسی داده‌هاست. در این فصل سعی می‌کنیم کمیّت‌هایی را معرفی کنیم که بتوانند به‌صورت کمّی، جامعه را معرفی نمایند. در رسم نمودارهای چندبر فراوانی و یا مستطیلی توجه کرده‌اید که معمولاً فراوانی داده‌ها در نقطه‌ای بیش‌تر از سایر نقاط است. به نمودارهای زیرتوجه کنید :



در نمودارهای بالا در فاصله‌های هاشوردار فراوانی از سایر قسمت‌ها بیش‌تر است. در واقع به‌نظر می‌رسد که داده‌ها بیش‌تر در این قسمت متمرکز شده‌اند. ما می‌خواهیم شاخصی را معرفی کنیم که محل تمرکز داده‌ها را معرفی کند. از این‌رو این شاخص‌ها را **شاخص‌های مرکزی** می‌گوییم. البته تاکنون از این قبیل شاخص‌ها بدون آن‌که معرفی شده باشند استفاده می‌کردیم، مثلاً :

ــ اگر بخواهند وضعیت تحصیلی شما را بدانند، اولین اطلاعی که از شما می‌خواهند، معدل نمرات شماست.

ــ اگر بخواهند میزان علاقهٔ شما را دربارهٔ زمینه‌های مختلف علمی بدانند از شما دربارهٔ تعداد کتاب‌هایی که در زمینه‌های مختلف مطالعه کرده‌اید، سؤال می‌شود. مثلاً می‌پرسند در چه زمینه‌ای بیش‌ترین تعداد کتاب را مطالعه کرده‌اید؟

ـ وضعیت تحصیلی شما در کلاس چگونه است؟ چه تعدادی از دانش‌آموزان کلاس را از نظر تحصیلی بهتر از خود ارزیابی می‌کنید. از چند نفر از دانش‌آموزان کلاس خود از نظر تحصیلی بهتر هستید؟ کدام‌یک از این دو عدد بزرگ‌تر است؟ مایلید کدام عدد بزرگ‌تر باشد؟

ملاحظه می‌کنید که این سؤال‌ها هر کدام به نوعی وضع تحصیلی شما را معرفی می‌کنند. ما در زیر به معرفی شاخص‌های نهفته در این سؤال‌ها می‌پردازیم.

مهم‌ترین شاخص‌های مرکزی عبارتند از : مد، میانه و میانگین، که به معرفی آن‌ها می‌پردازیم.



مد

یک کارخانه اتومبیل‌سازی می‌خواهد برای کشوری اتومبیل‌های سواری تولید کند. مدیر کارخانه می‌خواهد بداند کدام نوع اتومبیل مشتری‌های بیش‌تری دارد.

اتومبیل‌های با ظرفیت ۲ نفر

اتومبیل‌های با ظرفیت ۴ نفر

اتومبیل‌های با ظرفیت ۶ نفر

مدیر این کارخانه برای آن‌که بتواند تصمیم بگیرد چاره‌ای جز نمونه‌گیری ندارد. برای این منظور در یک نمونه‌گیری، از ۱۱۴ اتومبیل در حال حرکت مشاهده شده است که :

۴۸ دستگاه دارای ۱ یا ۲ سرنشین بوده‌اند.

۳۸ دستگاه دارای ۳ یا ۴ سرنشین بوده‌اند.

۲۸ دستگاه دارای ۵ یا ۶ سرنشین بوده‌اند.

این اعداد و ارقام نشان می‌دهد که بیش‌تر اتومبیل‌ها کم سرنشین بوده‌اند. پس اتومبیل‌های با ظرفیت ۶ نفره در این کشور استفادهٔ کم‌تری دارد. در این مثال اگر ماکزیمم تعداد سرنشین‌ها حساب شود، شاخصی به دست می‌آید که از نظر اقتصادی ممکن است به‌صرفه نباشد زیرا به علت وجود تعدادی

اتومبیل‌های با سرنشین‌های ۵ یا ۶ نفر ماکزیمم ۶ به‌دست می‌آید و لذا مدیر کارخانه تصمیم به‌تولید اتومبیل‌های بزرگ می‌گیرد که در این صورت هم از نظر قیمت و هم از نظر مصرف و هم از نظر تولید آلودگی، کاری زیان‌بار انجام شده است. *ما برای مطالعات خود باید شاخص‌های مناسب را انتخاب کنیم.*

مد، داده‌ای است که بیش‌ترین فراوانی را دارد.

در مثال بالا، نوع سرنشین کم، بیش‌ترین فراوانی (۴۸) را دارد پس مد مثال بالا اتومبیل کوچک است.

فروشندگان پوشاک از شاخص مد بسیار استفاده می‌کنند. آن‌ها با بررسی‌های خود درمی‌یابند که چه نوع پوشاکی مورد پسند مصرف‌کنندگان است. از همان پوشاک برای فروش سفارش می‌دهند، چون در شهریور ماه خرید لوازم‌التحریر خواهان بیش‌تری دارد، از این نوع اجناس بیش‌تر تولید و عرضه می‌کنند. در اعیاد بیش‌تر به تولید شیرینی و پوشاک و سایر ملزومات می‌پردازند. در نزدیکی عید، فروشنده نگران مشتری‌هایی که ممکن است احتیاج به کالایی داشته باشند که در این روزها مصرف چندانی ندارد، نیست.

مد ممکن است منحصر به فرد نباشد. مثلاً داده‌های

۵، ۵، ۷، ۵، ۲، ۱، ۴، ۳، ۲،۱، ۱، ۱

دارای دو مد ۵ و ۱ است. این قبیل جامعه‌ها را ۲ مدی می‌گویند. جامعه ممکن است چندمدی هم باشد. مد در این قبیل جامعه‌ها شاخص معتبری نیست. زیرا وجود چندین مد، نشان‌دهندهٔ این واقعیت است که جامعهٔ ما یک‌دست نیست (یادآوری می‌کنیم ما جامعه‌های یک‌دست را بررسی می‌کنیم) و لذا اگر مطالعات را دقیق‌تر کنیم، شاید بتوانیم جامعه را به دو بخش یک‌دست تفکیک کنیم. مثلاً ممکن است در بررسی قد افراد به ۲ مد برخورد کنیم که ممکن است نشان‌دهندهٔ این حقیقت باشد که ما در نمونه‌گیری بزرگ‌سالان و خردسالان را تفکیک نکرده‌ایم و لذا جامعه‌ای یک‌دست نداریم.

در رأی‌گیری‌ها، اساس تصمیم‌گیری مد است چون موضوعی که بیش‌ترین فراوانی را داشته باشد، انتخاب می‌شود. در انتخاب رئیس‌جمهور، نامزدی انتخاب می‌شود که بیش‌ترین فراوانی (رأی) را داشته باشد.

برای محاسبه مد فقط کافی است فراوانی داده‌ها را با هم مقایسه کنیم و دادهٔ با بیش‌ترین فراوانی مد است.



میانه

قضاوت شما دربارهٔ کلاسی که نصف دانش‌آموزان نمرهٔ بالای ۱۷ دارند چیست؟ اگر نمرهٔ نصف دانش‌آموزان بالای ۱۳ باشد قضاوت شما چگونه خواهد بود؟ در حالت اول خواهید گفت کلاس، کلاس خوبی است و دانش‌آموزان و دبیران در آن خوب کار کرده‌اند. حتی نصف دیگر کلاس که نمرهٔ کم‌تر از ۱۷ دارند ممکن است تعدادی از نمره‌ها ۱۶، ۱۵ و غیره باشند که باز هم نشان دهندهٔ وضع خوب کلاس است. اما در حالت دوم نصف دانش‌آموزان نمرهٔ زیر ۱۳ دارند که این نشانهٔ ضعف در کلاس است و لذا باید دربارهٔ آن چاره‌اندیشی کرد. نمرهٔ ۱۷ در حالت اول و نمرهٔ ۱۳ در حالت دوم را **میانه** می‌گوییم. میانه یکی از شاخص‌های مرکزی است که می‌توان دربارهٔ وضعیت جامعه از آن استفاده کرد.

پس از مرتب کردن داده‌ها، مقداری را که تعداد داده‌های بعد از آن با تعداد داده‌های قبل از آن برابر است، میانه می‌نامیم.

آیا لزوماً داده‌ها باید به‌صورت صعودی مرتب شوند؟ آیا می‌توان داده‌ها را به‌صورت نزولی نیز مرتب کرد؟

تعداد روزهایی که ۵ خانواده در تعطیلات عید به مسافرت رفته‌اند در زیر آمده است.

۳ ۴ ۶ ۷ ۱

می‌خواهیم میانه این داده‌ها را محاسبه کنیم. پس باید داده‌ها را مقایسه کنیم و لازم است آن‌ها را مرتب کنیم.

۷ ۶ ۴ ۳ ۱

حال ملاحظه می‌شود که عدد ۴ در بالا دارای ویژگی زیر است :

تعداد داده‌های بعد از ۴ ۲ تا

تعداد داده‌های قبل از ۴ ۲ تا

این دو تعداد با هم برابرند لذا، میانه داده‌های بالا برابر ۴ است. می‌بینید که در این مثال نصف تعداد داده‌ها که ۲/۵ است معنی ندارد.

آیا میانه همواره عضوی از مجموعه داده‌ها است؟ آیا می‌توانید مثالی بزنید که میانه عضوی از مجموعه داده‌ها نباشد؟ در زیر ساعات مطالعه آزاد دانش‌آموزان در طول هفته آمده است.

مثال: میانهٔ داده‌های زیر را تعیین کنید.

۱۰ ۴ ۳ ۹ ۶ ۲





برای محاسبهٔ میانه ابتدا داده‌ها را مرتب می‌کنیم.

۲ ۳ ۴ ۶ ۹ ۱۰

می‌بینیم که هیچ یک از داده‌ها ویژگی میانه را ندارد ولی اگر شما عددی بین ۴ و ۶ مثلاً ۴/۳ را در نظر بگیرید، این عدد در تعریف میانه صدق می‌کند زیرا تعداد داده‌های بعد از ۴/۳ برابر ۳ و تعداد داده‌های قبل از ۴/۳ برابر با ۳ است. از این‌جا نتیجه می‌شود که شما هر عددی را بین ۴ و ۶ در نظر بگیرید در تعریف میانه صدق می‌کند. لکن برای آن‌که یکسان عمل نماییم در این حالت میانه را به‌صورت زیر به دست می‌آوریم :

$$\frac{۴+۶}{۲}=۵$$

روش پیدا کردن میانه

۱ــ داده‌ها را مرتب کنیم.

۲ــ اگر تعداد داده‌ها فرد باشد، داده‌ای که در وسط قرار می‌گیرد برابر میانه است.

۳ــ اگر تعداد داده‌ها زوج باشد، نصف مجموع دو داده‌ای که در وسط قرار گرفته‌اند برابر میانه است.

دو دسته داده‌های زیر را در نظر بگیرید

۰ ,۲ ,۱۰۰ ۱ ,۲ ,۳

هر دو دسته داده دارای میانه ۲ هستند ولی ملاحظه می‌کنید که تفاوت‌های اساسی بین آن‌ها وجود دارد. *میانه فقط به این نکته توجه دارد که تعداد داده‌های بعد از آن برابر تعداد داده‌های قبل از آن باشد و به بزرگی و یا کوچکی داده‌ها کاری ندارد. به عبارت دیگر، میانه نسبت به اندازهٔ داده‌ها حساسیت نشان نمی‌دهد.*

اگر میانه از مد کوچک‌تر باشد چگونه داده‌ها را تحلیل می‌کنید؟

اگر میانه از مد بزرگ‌تر باشد چه اطلاعاتی در مورد داده‌ها می‌دهد؟

آیا مد نسبت به اندازهٔ داده‌ها حساسیت نشان می‌دهد؟

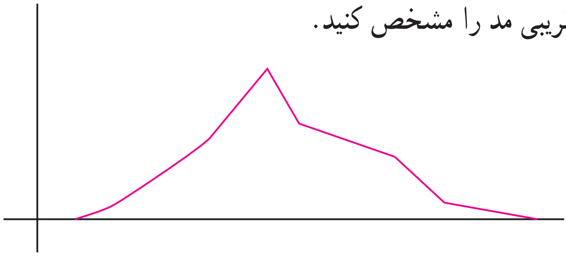
تمرین

۱ــ میانه و مد را در داده‌های زیر حساب کنید.

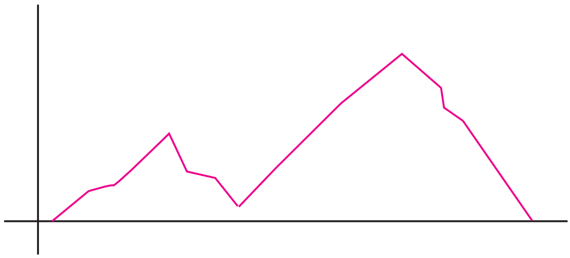
۰، ۵، ۶، ۵، ۱۵، ۱۲، ۰، ۴، ۱۵، ۵

۲ــ اگر داده‌ها را دو برابر کنیم چه تأثیری در مقدار میانه و مد حاصل می‌شود؟

۳ــ در نمودار زیر محل تقریبی مد را مشخص کنید.



۴ ــ اگر در مطالعه قد، نمودار به شکل زیر باشد چه نتیجه‌ای می‌گیرید؟





نمودار جعبه‌ای

نمودارهایی که تاکنون شناخته‌ایم هرکدام به طریقی داده‌ها را نمایش می‌دادند. فکر می‌کنید کدام نمودار بهتر از بقیه پراکندگی داده‌ها را نشان می‌دهد؟

موارد استفادهٔ هر یک از انواع نمودارها را بیان کنید.

همان‌طور که می‌بینید این نمودارها برای مقایسه داده‌ها بسیار مفیدند ولی هیچ‌کدام به سؤالاتی از قبیل : آیا داده‌ها به هم نزدیک هستند؟ آیا داده‌ها بیش‌تر در اطراف میانگین متمرکزند یا بیش‌تر اطراف کم‌ترین داده یا بیش‌ترین داده متمرکزند؟ پاسخ نمی‌دهند.

فعالیت ۱

تعداد تصادفات اتومبیل در شهری در ۱۵ روز اول تابستان عبارتند از :

۱۶، ۲۷، ۱۴، ۲۳، ۱۵، ۱۰، ۱۲

۱۹، ۳۱، ۲۵، ۱۸، ۳۲، ۴۳، ۴۱، ۳۴

ــ میانه را پیدا کنید.

ــ اعدادی را که در نیمهٔ قبل از میانه قرار دارند بنویسید.

ــ برای این اعداد، میانه را پیدا کنید.

ــ اعدادی را که در نیمهٔ بعد از میانه قرار دارند بنویسید.

ــ برای این اعداد میانه را پیدا کنید.

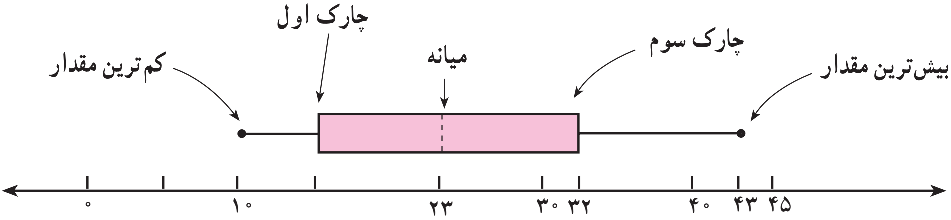
میانهٔ نیمهٔ اول داده‌ها را چارک اول Q_1 و میانهٔ نیمهٔ دوم داده‌ها را چارک سوم Q_3 می‌نامند.

ــ چارک دوم یا Q_2 چه نام دارد؟

نمودار جعبه‌ای نموداری تصویری است که داده‌ها را براساس پنج مقدار نمایش می‌دهد. این مقادیر عبارتند از :

۱ــ کوچک‌ترین داده　　۲ــ چارک اول　　۳ــ میانه

۴ــ چارک سوم　　۵ــ بزرگ‌ترین داده

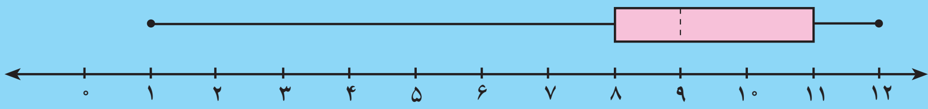


نمودار فوق نمودار جعبه‌ای داده‌های موجود در فعالیت ۱ می‌باشد.

ــ چند درصد از داده‌ها در داخل جعبه قرار دارد؟

ــ چند درصد از داده‌ها در هر یک از دنباله‌ها قرار دارند؟

فعالیت ۲



ــ در نمودار بالا میانه چند است؟

ــ چارک اول در این داده‌ها ۸ است، این عدد نشان‌دهندهٔ چیست؟

ــ چارک سوم چند است؟ این عدد نشان‌دهندهٔ چیست؟

ــ وجود میانه در سمت چپ جعبه نشان‌دهندهٔ چیست؟

ــ بلندتر بودن دنبالهٔ سمت چپ نسبت به دنبالهٔ سمت راست نشان‌دهندهٔ چیست؟

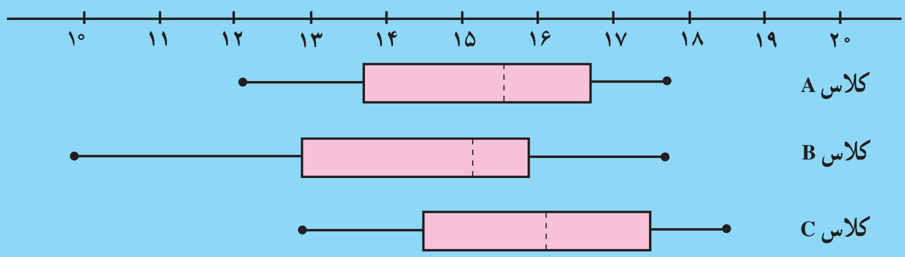




فعالیت ۳

نمودار زیر عمل‌کرد سه کلاس A ، B و C را در امتحان ریاضی نشان می‌دهد.

با توجه به نمودار به سؤالات زیر پاسخ دهید.



ــ کدام کلاس بیش‌ترین دامنهٔ تغییرات را دارد؟

ــ میانهٔ نمرات کدام کلاس از همه بیش‌تر است؟ میانهٔ نمرات کدام کلاس از همه کم‌تر است؟

ــ پراکندگی نمرات کدام کلاس بیش‌تر از همه است؟

ــ این سه کلاس را با توجه به عمل‌کردشان در امتحان، از ضعیف‌ترین به قوی‌ترین مرتب کنید.

فعالیت ۴

آمار در پزشکی و بیمه کاربردهای بسیاری دارد. به‌طور مثال یک شرکت بیمه می‌خواهد از آماری که حاوی اطلاعاتی در مورد طول مدت بستری شدن پس از سکتهٔ قلبی در بیمارستان است استفاده کند و روش‌ها و برنامه‌های خود را برای بیمه نمودن بیماران مورد بررسی قرار دهد. داده‌ها به شرح زیر می‌باشند :

تعداد بیماران زن	تعداد بیماران مرد	طول مدت بستری شدن پس از سکته قلبی (روز)
۴	۶	۰
۲	۱۵	۱
۳	۴	۲
۴	۲	۳
۲	۵	۴
۱	۸	۵
۴	۱۶	۶
۲	۱۲	۷
۵	۹	۸
۵	۱۱	۹
۰	۱۰	۱۰
۰	۱۸	۱۱
۲	۷	۱۲
۳	۲	۱۳
۲	۵	۱۴
۰	۷	۱۵
۲	۱	۱۶
۰	۳	۱۷
۱	۱	۱۸
۰	۱	۱۹
۰	۱	۲۱
۱	۲	۲۲
۱	۰	۲۳
۱	۱	۲۴
۰	۱	۲۶
۱	۱	۲۸
۲	۰	۲۹
۰	۲	۳۷
۰	۱	۴۱
۱	۰	۴۵
۱	۰	۷۲

ــ نمودار جعبه‌ای را برای مدت بستری شدن زنان و مردان به‌طور جداگانه و برای کل بیماران رسم کنید. این نمودارها را مقایسه کنید و در کلاس به بحث بگذارید.

تمرین

۱ــ جهت تعیین مقاومت یک حیوان آزمایشگاهی در مقابل بیماری سل (باسیل توبرکلوسیس)، به ۷۳ حیوان این باسیل تزریق شد، پس از ۴۳ روز اولین حیوان مُرد. قوی‌ترین آن‌ها پس از ۵۹۸ روز مُرد. جدول زیر طول مقاومت این حیوان را در مقابل این باسیل نشان می‌دهد :

۴۳, ۴۵, ۵۳, ۵۶, ۵۶, ۵۷, ۵۸, ۶۶, ۶۷, ۷۳

۷۴, ۷۹, ۸۰, ۸۰, ۸۱, ۸۱, ۸۱, ۸۲, ۸۳, ۸۳

۸۴, ۸۸, ۹۱, ۹۱, ۹۱,۹۲, ۹۲, ۹۷, ۹۹, ۹۹, ۱۰۰

۱۰۰, ۱۰۲, ۱۰۲, ۱۰۲, ۱۰۲, ۱۰۳, ۱۰۴, ۱۰۷

۱۰۸, ۱۰۹, ۱۱۳, ۱۱۸, ۱۱۸, ۱۲۱, ۱۲۳, ۱۲۶

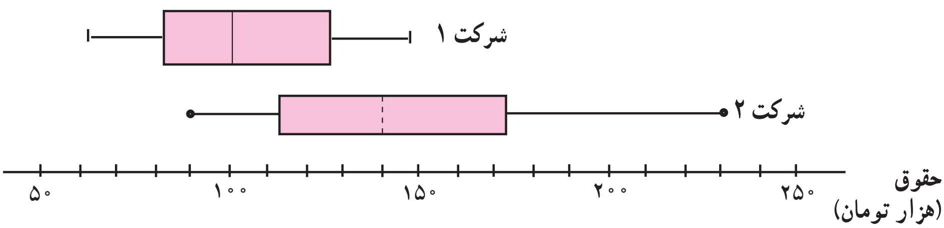
۱۲۸, ۱۳۷, ۱۳۸, ۱۳۹, ۱۴۴, ۱۴۷, ۱۴۷, ۱۵۶

۱۶۲, ۱۷۴, ۱۷۸, ۱۷۹, ۱۸۴, ۱۹۱, ۱۹۸, ۲۱۴, ۲۱۴

۲۴۳, ۲۴۹, ۲۴۹, ۳۲۹, ۳۸۰, ۴۰۳, ۵۱۱, ۵۲۲, ۵۹۸

آیا نمودار ساقه و برگ برای توصیف این داده‌ها، نمودار مناسبی است؟ چرا؟ نمودار جعبه‌ای این داده‌ها را رسم کنید.

۲ــ نمودار زیر نشان‌دهندهٔ پرداخت حقوق کارمندان در دو شرکت می‌باشد.



«حقوق تقریباً ۵۰٪ کارمندان شرکت ۲ بیش‌تر از ماکزیمم حقوق کارمندان شرکت ۱ است.»

«حقوق تمام کارمندان شرکت ۲ بیش‌تر از حقوق ۵۰٪ از کارمندان شرکت ۱ است.»

ــ آیا ادعاهای فوق با نمودار مطابقت دارند؟

ــ از چنین نموداری چه استفاده‌هایی می‌توان کرد؟



میانگین

شما تاکنون معدل خود را بارها حساب کرده‌اید، برای این‌کار نمرات را جمع کرده‌اید و بر تعداد آن‌ها تقسیم کرده‌اید. در واقع میانگین یعنی معدل داده‌ها. به‌طور صریح فرض کنید داده‌های ما به‌صورت :

$$x_1, x_1, \ldots, x_n$$

باشد، در این صورت میانگین آن‌ها که با نماد $\bar{x}$ نشان می‌دهیم عبارت است از :

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}$$

فرض کنید میانگین نمرات ۷ درس شما ۱۵ شده باشد و میانگین نمرات همان ۷ درس برای یکی از همکلاسی‌های شما برابر ۱۴ باشد. پس مجموع نمرات شما در این هفت درس برابر ۱۰۵ = ۷×۱۵ و مجموع نمرات همکلاسی شما برابر ۹۸ = ۷×۱۴ است. یعنی اگر برای هر درسی



امتیازی بین ۰ تا ۲۰ (که همان نمرهٔ درس است) بدهند شما از این ۷ درس ۱۰۵ امتیاز و همکلاسی شما ۹۸ امتیاز آورده است. این دو امتیاز نشان می‌دهند که وضع تحصیلی شما از وضع تحصیلی همکلاسی شما بهتر است. اگر تعداد درس‌ها زیاد باشد، جمع امتیازها عدد بزرگی خواهد شد و مقایسه و تحلیل آن‌ها به‌راحتی در ذهن انجام نمی‌شود.

ــ اگر تعداد درس‌ها برابر نباشند، مثلاً شما ۱۱۰ امتیاز در ۸ درس و همکلاسی شما ۹۰ امتیاز در ۶ درس آورده باشد، کدام‌یک از شما دو نفر بهتر از دیگری عمل کرده‌اید؟

ظاهراً امتیاز شما بیش‌تر است ولی این دلیل بهتر بودن شما نیست، مسلماً اگر تعداد درس‌ها را بیش‌تر کنیم، امتیاز بالاتر خواهد رفت. برای آن‌که بتوانیم این مقایسه را انجام دهیم، باید ببینیم که اگر قرار باشد در تمام درس‌ها امتیازات برابر آورده باشید به قسمتی که جمع امتیاز ثابت باشد، آن امتیاز کدام است. مثلاً اگر برای هر یک از ۸ درس شما امتیاز a فرض کرده باشیم، جمع امتیازهای شما برابر $a+a+\cdots+a=8a$ خواهد بود که این مقدار باید برابر ۱۱۰ شود. پس امتیاز مشترک درس‌های شما برابر $\frac{110}{8}=13/75$ که همان میانگین نمرات شماست به‌دست می‌آید. امتیاز مشترک درس‌های همکلاسی شما برابر $\frac{90}{6}=15$ است. این اعداد و ارقام نشان می‌دهند عموماً همکلاسی شما در درس‌های خود بهتر از شما عمل کرده است. نمرات شما بیش‌تر در اطراف ۱۳/۷۵ متمرکز شده‌اند و حال آن‌که نمرات همکلاسی شما در اطراف ۱۵ متمرکز شده‌اند. پس سطح نمرات همکلاسی شما بالاتر است.

جایی در توضیحات بالا گفتیم که نمرات شما در اطراف ۱۳/۷۵ متمرکز شده است، میانگین در مرکز داده قرار دارد، یعنی همان مقداری که داده‌ها از میانگین اضافی دارند، همان مقدار هم از میانگین کم دارند. این موضوع در قضیهٔ صفحهٔ بعد بیان شده‌است.

قضیه: اگر $\bar{x}$ ، میانگین $x_1, \cdots, x_n$ باشد آن‌گاه مجموع اختلافات داده‌ها از میانگین برابر است با صفر.

اثبات:

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + \cdots + (x_n - \bar{x})$$

$$= x_1 + x_2 + \cdots + x_n - \underbrace{\bar{x} - \bar{x} - \cdots - \bar{x}}_{\text{n تا}} = x_1 + x_2 + \cdots + x_n - (\underbrace{\bar{x} + \bar{x} + \cdots + \bar{x}}_{\text{n تا}}) = 0$$

زیرا می‌دانیم مجموع مقادیر برابر حاصل‌ضرب میانگین در تعداد آن‌ها است یعنی :

$x_1 + x_2 + \cdots + x_n = n\bar{x}$ □

فعالیت

جدول زیر قد ۳۰ دانش‌آموز را برحسب سانتی‌متر نشان می‌دهد :

۱۶۰، ۱۵۰، ۱۵۵، ۱۳۷، ۱۷۵، ۱۷۷

۱۷۷، ۱۶۹، ۱۴۰، ۱۷۹، ۱۴۶، ۱۳۲

۱۳۰، ۱۳۵، ۱۷۶، ۱۳۰، ۱۴۸، ۱۴۱

۱۳۶، ۱۴۳، ۱۴۳، ۱۷۸، ۱۷۲، ۱۵۷

۱۷۸، ۱۴۵، ۱۵۱، ۱۳۰، ۱۶۶، ۱۳۲

ـ میانگین قد این دانش‌آموزان را حساب کنید.

ـ اگر به تمام این داده‌ها ۵ تا اضافه شود، چه تغییری در میانگین حاصل می‌شود؟

ـ اگر از تمام این داده‌ها ۱۲ تا کم شود، چه تغییری در میانگین حاصل می‌شود؟

ـ با اضافه کردن a چه تغییری در میانگین حاصل می‌شود؟

ـ اگر تمام داده‌ها در ۲ ضرب شوند، چه تغییری در میانگین حاصل می‌شود؟

ـ اگر تمام داده‌ها در ۰/۲ ضرب شوند، چه تغییری در میانگین حاصل می‌شود؟

ـ اگر تمام داده‌ها در $a > 0$ ضرب شوند چه تغییری در میانگین حاصل می‌شود؟

ـ اگر تمام داده‌ها در $a < 0$ ضرب شوند، چه تغییری در میانگین حاصل می‌شود؟

ـ با توجه به نتایج به‌دست آمده در بالا، میانگین داده‌های زیر را به‌دست آورید.

الف ـ ۱۴۱۲، ۱۳۰۰، ۱۳۴۲، ۱۳۰۷، ۱۳۱۷، ۱۳۲۸

ب ـ ۸۰۳۸، ۱۰۰۳۸، ۳۰۳۸، ۴۰۳۸، ۱۰۳۸، ۲۰۳۸

در برخی از محاسبات ممکن است اعداد بسیار بزرگ باشند، جمع کردن این اعداد و تقسیم آن‌ها همراه با خطاهایی خواهد بود. شاید بگویید در عصر رایانه‌ها، بزرگی اعداد مشکلی نخواهد بود. ولی به‌هر حال این اعداد را باید به طریقی وارد کنیم که در این مرحله بزرگی آن‌ها، اشکالاتی را ایجاد خواهد کرد. اگر بتوانیم تدبیری بیندیشیم تا از اعداد کوچک‌تر استفاده کنیم، این مشکل را دور زده‌ایم. قضیه زیر راه‌حل مناسب این مسئله را ارائه می‌کند.

قضیه: اگر $\bar{x}$ میانگین داده‌های $x_۱,\dots,x_n$ باشد آنگاه میانگین داده‌های $ax_۱+b,\dots,ax_n+b$ که با نماد $\overline{ax+b}$ نشان می‌دهیم عبارت است از :

$$\overline{ax+b}=a\bar{x}+b$$

اثبات:

$$\overline{ax+b}=\frac{(ax_۱+b)+\cdots+(ax_n+b)}{n}$$
$$=\frac{a(x_۱+\cdots+x_n)+(b+\cdots+b)}{n}$$
$$=a\frac{x_۱+\cdots+x_n}{n}+\frac{nb}{n}$$
$$=a\bar{x}+b$$

مثال: میانگین داده‌های زیر را حساب کنید.

۱۷۵۰ , ۱۷۶۰ , ۱۷۶۶ , ۱۷۵۵ , ۱۷۸۰

حل: ملاحظه می‌کنیم که اعداد بالا به‌صورت زیر هستند.

$۱۷۰۰+۱۰\times ۵, ۱۷۰۰+۱۰\times ۶, ۱۷۰۰+۱۰\times ۶/۶, ۱۷۰۰+۱۰\times ۵/۵, ۱۷۰۰+۱۰\times ۸$

پس $a=۱۰$ و $b=۱۷۰۰$ و کافی است میانگین داده‌ها

۵، ۶، ۶/۶، ۵/۵، ۸

را حساب کنیم. میانگین این اعداد عبارت است از :

$$\frac{۵+۶+۶/۶+۵/۵+۸}{۵}=\frac{۳۱/۱}{۵}=۶/۲۲$$

پس میانگین داده‌های اصلی عبارت است از :

$$۱۷۰۰+۱۰\times ۶/۲۲=۱۷۰۰+۶۲/۲=۱۷۶۲/۲$$

مقایسه میانگین و میانه: میانگین شاخص خوبی برای نشان دادن مرکزیت داده‌هاست. در بعضی از مسائل این شاخص نمی‌تواند مؤثر باشد. مثلاً انتظار می‌رود در یک کلاس تمامی دانش‌آموزان نمرهٔ خوبی بگیرند. اگر نمرات خوب باشد مسلماً میانگین هم بالا خواهد بود ولی آیا بالا بودن میانگین

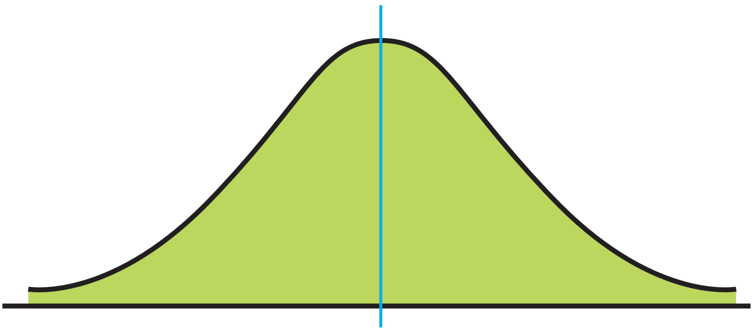
نمرات نشان دهندهٔ وضع خوب کلاس است. ممکن است تعداد کمی نمرهٔ ۲۰ گرفته باشند و تعداد زیادی نمره ۱۰ یا ۱۲. وضع این کلاس رضایت‌بخش نخواهد بود برای آن‌که وضع کلاس را بتوانیم خوب ارزیابی کنیم حداقل نصف کلاس باید نمرهٔ خوبی آورده باشند.

آن چه نمره‌ای است که نمرهٔ نصف دانش‌آموزان کلاس از آن بیش‌تر است؟

ــ اگر میانه خیلی از میانگین کوچک‌تر باشد تعبیر آن چیست؟

ــ اگر میانه خیلی بزرگ‌تر از میانگین باشد تعبیر آن چیست؟

در فصل نمودارها به منحنی نرمال اشاره کردیم و گفتیم که این یک منحنی متقارن زنگی شکل به‌صورت زیر است.



از تقارن آن نتیجه می‌شود که میانه و میانگین آن برابرند و چون نقطه ماکزیمم هم دارند مد آن نیز برابر میانگین است.



میانگین وزن‌دار

شما قطعاً شنیده‌اید که برخی از نمرات شما ضریب دارند، مثلاً اگر چهار بار از شما آزمون گرفته باشند، نمره آخرین آزمون را ممکن است در ۳ ضرب کنند، یعنی به آن ضریب ۳ بدهند. پس اگر نمرات شما به‌صورت زیر باشد :

آزمون چهارم	آزمون سوم	آزمون دوم	آزمون اول
۱۷	۱۸	۱۵	۱۶

نمره‌ای که برای این درس برای شما در کارنامه خواهد آمد میانگین ۶ عدد زیر است :

۱۷ ۱۷ ۱۷ ۱۸ ۱۵ ۱۶

یعنی نمره درس مذکور عبارت است از :

$$\frac{16+15+18+17+17+17}{6}=\frac{100}{6}\approx 16/66$$

ابتدا ببینیم این ضریب چه اثری در نمرهٔ شما می‌گذارد. نمره بدون احتساب ضریب عبارت است از :

$$\frac{16+15+18+17}{4}=16/5$$

اگر ضریب را بیش‌تر کنیم مثلاً نمرهٔ آخر را با ضریب ۵ در نظر بگیریم خواهیم داشت :

$$\frac{16+15+18+17\times 5}{8}=16/75$$

ملاحظه می‌شود که هر چه‌قدر ضریب نمره آزمون چهارم بالاتر رود نمرهٔ درس شما بیش‌تر خواهد شد. حال ببینیم آیا همیشه این‌طور است. فرض کنید نمرات شما به‌صورت زیر باشد :

۱۶ ۱۸ ۱۵ ۱۶

در این‌صورت نمرهٔ درس شما بدون احتساب ضریب عبارت است از :

$$\frac{16+15+18+16}{4}=16/25$$

حال اگر به نمرهٔ آزمون چهارم ضریب ۳ بدهیم، خواهیم داشت :

$$\frac{16+15+18+16+16+16}{6}\approx 16/16$$

اگر ضریب آزمون چهارم ۵ باشد نمره درس شما عبارت است از :

$$\frac{16+15+18+5\times 16}{8}=16/125$$

این‌بار ملاحظه می‌کنید که هر چه‌قدر ضریب بالاتر می‌رود نمرهٔ شما کم‌تر می‌شود. می‌توانید علت آن را توضیح دهید؟



چگونه است که در مثال قبل با افزایش ضریب، نمره بالاتر می‌رفت ولی در این مثال با افزایش ضریب نمره کاهش می‌یابد؟

شاید دقت کرده باشید که در مثال اول نمرهٔ آزمون چهارم بیش از میانگین نمرات (بدون ضریب شماست. میانگین نمرات بدون ضریب برابر ۱۶/۵ بود ولی نمرهٔ آزمون چهارم شما ۱۷ بود. امّا در مثال دوم میانگین نمرات بدون ضریب برابر ۱۶/۲۵ و نمره آزمون چهارم شما ۱۶ بود. هر چه‌قدر این نمره از میانگین فاصلهٔ بیش‌تری داشته باشد تأثیر ضریب آن محسوس‌تر است. این مطالب را در جدول زیر به وضوح می‌بینید.

نمره چهارم با ضریب ۴	نمره چهارم با ضریب ۲	میانگین بدون ضریب	نمرات
۱۶/۱۴	۱۵/۸	۱۵/۵	۱۵،۱۶،۱۴،۱۷
۱۷/۲۸	۱۶/۶	۱۶	۱۵،۱۶،۱۴،۱۹
۱۳/۲۸	۱۳/۸	۱۴/۲۵	۱۵،۱۶،۱۴،۱۲

ملاحظه می‌کنید اگر نمره زیر میانگین باشد و ضریب آن بالاتر رود نمره درس کاهش می‌یابد برعکس اگر نمره بالاتر از میانگین باشد و ضریب آن بالاتر رود نمره درس افزایش می‌یابد. معمولاً

نمراتی که اهمیت بیش‌تری دارند باید بتوانند جایگاه خود را در بین نمرات نشان دهند. سه نمرهٔ اول چون از بخش‌های کوچک کتاب امتحان گرفته شده است، انتظار می‌رود نمرات خوبی باشند ولی نمرهٔ آزمون چهارم چون از تمام کتاب بوده و از سؤال‌های ترکیبی (سؤالاتی که چند مفهوم را در هم ادغام کرده باشند) استفاده شده است، اهمیت آن بیش‌تر است این اهمیت از طریق ضریب اعمال می‌شود.

حال به حالت کلی برمی‌گردیم، فرض کنیم برحسب ضروریاتی به هر آزمونی ضریب اختصاص داده شده باشد، مثلاً فرض کنید در چهار آزمون برگزار شده به‌ترتیب ضریب‌های ۱، ۲، ۳، ۳ اختصاص داده شده باشد و نمرات شما به‌ترتیب از آزمون اول تا چهارم برابر ۱۴، ۱۷، ۱۶، ۱۵ باشد، نمره این درس شما چه‌قدر خواهد بود؟ در واقع می‌خواهیم میانگین مقادیر زیر را حساب کنیم.

۱۵، ۱۵، ۱۵، ۱۶، ۱۶، ۱۶، ۱۷، ۱۷، ۱۴

پس نمره درس عبارت است از :

$$\frac{۱۴+(۱۷+۱۷)+(۱۶+۱۶+۱۶)+(۱۵+۱۵+۱۵)}{۱+۲+۳+۳}$$

$$=\frac{۱۴۱}{۹}\approx ۱۵/۶۶$$

این قبیل میانگین‌ها را **میانگین‌های وزن‌دار (وزنی)** می‌گویند.

در حالت کلی فرض کنید داده‌های $x_۱, \dots, x_n$ به‌ترتیب دارای ضریب‌های $w_۱, \dots, w_n$ باشد. این اطلاعات را برای سادگی می‌توانیم در جدول زیر خلاصه کنیم :

داده‌ها	$x_۱$	$x_۲$	$\cdots$	x_n
ضریب (وزن)	$w_۱$	$w_۲$	$\cdots$	w_n

در این صورت میانگین داده‌های بالا با احتساب ضرایب مربوطه به‌صورت زیر محاسبه می‌شود :

$$\text{میانگین وزنی} = \frac{w_۱x_۱+w_۲x_۲+\cdots+w_nx_n}{w_۱+w_۲+\cdots+w_n}$$

مثال: میانگین وزن‌دار را در جدول زیر حساب کنید.

حل:

داده	۱۲	۱۴/۵	۱۵	۹
وزن	۱	۳	۱	۳

$$\text{میانگین وزنی} = \frac{۱\times۱۲+۳\times۱۴/۵+۱\times۱۵+۳\times۹}{۱+۳+۱+۳}$$

$$=\frac{۹۷/۵}{۸}\approx ۱۲/۱۸$$

میانگین وزن‌دار را هم با همان نماد $\bar{x}$ نشان می‌دهند و اگر بخواهند تمایز بین میانگین معمولی و میانگین وزن‌دار قائل شوند، میانگین وزن‌دار را با نماد $\bar{x}_w$ نشان می‌دهند.

در مواردی که ما با میانگین وزن‌دار مواجه می‌شویم محاسبه میانگین در یک جدول فراوانی است. دیدیم که در یک جدول فراوانی قرار شد تمام افرادی که در یک دسته قرار دارند برابر مرکز دسته فرض کنیم. پس مرکز دسته به اندازهٔ فراوانی آن دسته تکرار می‌شود، و این مانند آن است که ما به مرکز دسته وزنی برابر فراوانی آن دسته داده باشیم، پس اگر بخواهیم میانگین را حساب کنیم باید مرکز دسته‌ها را در فراوانی‌های نظیر، ضرب و با هم جمع کرده و پس از آن بر مجموع فراوانی‌ها (که همان تعداد داده‌هاست) تقسیم کنیم.

میانگین در جدول فراوانی : $\bar{x} = \frac{f_۱x_۱ + \cdots + f_k x_k}{f_۱ + \cdots + f_k}$

این مفهوم میانگین برای محاسبات از روی جدول فراوانی مناسب‌تر است.

داده‌های زیر را که ضریب هوشی ۳۴ دانش‌آموز پسر در سال اول دبیرستان را نشان می‌دهد در نظر بگیرید.

۸۸	۱۰۱	۱۰۸	۱۰۳	۱۱۱	۱۱۱	۱۰۸	۱۰۸	۸۸
۹۶	۱۰۷	۱۱۱	۹۶	۱۱۲	۱۰۵	۱۰۵	۱۰۳	۹۱
۱۰۵	۹۰	۹۵	۸۹	۹۲	۱۱۲	۱۱۲	۱۰۰	۹۲
	۷۷	۹۰	۱۰۶	۷۷	۱۰۸	۹۵	۱۱۱	

ابتدا مستقیماً میانگین این داده‌ها را حساب کنید. سپس با تشکیل یک جدول فراوانی مناسب میانگین را به‌دست آورده و جواب‌ها را مقایسه کنید.

حل: میانگین داده‌های بالا عبارت است از :

$$\frac{\text{مجموع داده‌ها}}{۳۴} = \frac{۳۴۰۳}{۳۴} \approx ۱۰۰/۹$$

حال داده‌ها را در دسته‌هایی، دسته‌بندی می‌کنیم. ابتدا ملاحظه می‌کنیم که

$a = ۷۷ \ , \ b = ۱۱۲$

پس $R = b - a = ۳۵$. این ۳۵ داده را در ۵ دسته، دسته‌بندی می‌کنیم.

طول دسته‌ها $= C = \frac{۳۵}{۵} = ۷$

فراوانی‌ها	مرکز دسته‌ها	دسته‌ها
۲	۸۰/۵	۷۷ − ۸۴
۵	۸۷/۵	۸۴ − ۹۱
۷	۹۴/۵	۹۱ − ۹۸
۴	۱۰۱/۵	۹۸ − ۱۰۵
۱۶	۱۰۸/۵	۱۰۵ − ۱۱۲
۳۴	جمع	

$$\overline{x}_w=\frac{۲\times۸۰/۵+۵\times۸۷/۵+۷\times۹۴/۵+۴\times۱۰۱/۵+۱۶\times۱۰۷/۵}{۳۴}$$

$$=\frac{۳۳۸۶}{۳۴}\approx۹۹/۵۸$$

ملاحظه می‌کنید آن‌چه که از روی جدول فراوانی به‌دست آمد با آن‌چه که مستقیماً از روی داده‌ها به‌دست آمد به اندازهٔ $۱۰۰/۰۹-۹۹/۵۸=۰/۵۱$ اختلاف دارد. این اختلاف در مقایسه با میانگین داده‌ها که حدوداً برابر ۱۰۰ است ناچیز است. قبلاً هم گفتیم در جدول‌بندی و یکسان‌سازی داده‌ها مقداری از اطلاعات از دست می‌رود ولی سهولت‌ها و تعبیرهایی که این روش برای ما ایجاد خواهد کرد به‌مراتب باارزش‌تر از حفظ مقدار ناچیزی از اطلاعات است.

توجه ۱: اگر بخواهیم برابری $(x_۱-\overline{x})+\ ...\ +(x_n-\overline{x})=۰$ را در جدول فراوانی بررسی کنیم باید فراوانی‌ها را نیز در نظر بگیریم. یعنی در جدول فراوانی داریم :

$$f_۱(x_۱-\overline{x})+f_۲(x_۲-\overline{x})+\ ...\ +f_n(x_n-\overline{x})=۰$$

در مثال بالا داریم :

$$۲(۸۰/۵-۱۰۰/۱)+۵(۸۷/۵-۱۰۰/۱)+$$
$$۷(۹۴/۵-۱۰۰/۱)+۴(۱۰۱/۵-۱۰۰/۱)+$$
$$۱۶(۱۰۸/۵-۱۰۰/۱)$$
$$=-۳۹/۲-۶۳-۳۹/۲+۵/۶+۱۳۴/۴=۱/۴$$

ــ البته این مقدار صفر نیست ولی به صفر بسیار نزدیک است آیا می‌توانید توضیح دهید که چرا دقیقاً برابر صفر به‌دست نیامد؟

روشی سریع برای محاسبهٔ میانگین: سعید نمرات زیر را در پایان نیمسال اوّل اتخاذ کرده است. او برای محاسبه سریع معدل نمرات خود، ابتدا معدل خود را حدس می‌زند مثلاً فرض می‌کند معدل او ۱۸ است. سپس از نمره اوّل شروع می‌کند و با معدل حدسی خود مقایسه می‌کند.

لیست نمرات	۱۵	۱۷	۱۵	۱۹	۲۰	۱۹	۱۳	۱۶	۱۸	۱۹
تفاضل از معدل تخمینی ۱۸	۳−	۱−	۳−	۱+	۲+	۱+	۵−	۲−	۰	۱+
جمع مقادیر مقایسه شده	۳−	(۱−)+۳− ۴−	(۳−)+۴− ۷−	=۱+۷− ۶−	۴−	۳−	۸−	۱۰−	۱۰−	۹−

نمره اوّل ۱۵ است که نسبت به معدل تخمینی ۳،۱۸ نمره کم دارد، این تفاوت را با ۳- نشان می‌دهیم. حال نمره دوم را در نظر می‌گیرد. نسبت به معدل ۱۸ نمرهٔ دوم ۱ نمره کم است. او تاکنون ۴ نمره از معدل حدسی خود کم آورده است. همین روش را ادامه می‌دهد و در پایان او ۹ نمره از معدل کم آورده است. ۹ را بر تعداد نمرات تقسیم می‌کند و مقدار آن را از میانگین حدسی کم می‌کند، تا میانگین واقعی به‌دست آید.

$$\bar{x} = ۱۸ - \frac{۹}{۱۰} = ۱۸ - ۰/۹$$

$$= ۱۸ - ۰/۹ = ۱۷/۱$$

در حالت کلی فرض کنید $x_۱, \ldots, x_n$ داده‌ها باشند و y میانگین تخمینی باشد.

$$A = \frac{(x_۱ - y) + \cdots + (x_n - y)}{n}$$

$$= \frac{x_۱ + \cdots + x_n - ny}{n}$$

$$= \bar{x} - y$$

پس :

$$\bar{x} = y + A$$

در مثال مقدار A را برابر $\frac{-۹}{۱۰}$ حساب کردیم و از رابطه $A = \bar{x} - y$ با داشتن A و y میانگین واقعی $\bar{x}$ را به‌دست آوردیم

فردی را شکنجه می‌دادند. شکنجهٔ این فرد به این صورت بود که یکی از دست‌های او را در کوره‌ای با دمای ۱۰۰ درجه سانتی‌گراد بالای صفر قرار دادند و دست دیگر او را در داخل مبرّدی با سرمای ۱۰۰- درجه سانتی‌گراد قرار دادند. در این حال از او پرسیدند، در چه حالی هستی ؟ او در پاسخ گفت : به‌طور متوسط احساس راحتی می‌کنم.

تمرین

۱ــ بهزاد در آزمایشگاه فیزیک، پس از انجام یک آزمایش به داده‌های زیر دست پیدا کرد :

۲۰۲۰، ۲۰۹۶، ۲۲۰۲، ۲۰۱۶، ۲۰۷۳، ۲۰۳۹

ــ میانگین و میانه و مد را برای داده‌های به‌دست آمده به‌دست آورید.

ــ پس از بررسی مجدد داده‌ها، بهزاد متوجه شد که به‌جای ۲۲۰۲ باید ۲۰۲۲ ثبت می‌شد. او این تغییر را در لیست خود اعمال نمود.

ــ میانگین، میانه و مد را برای لیست جدید داده‌ها به‌دست آورید.

ــ کدام‌یک از شاخص‌های فوق با این تغییر، بیش‌تر از همه تغییر یافت؟

ــ عموماً کدام شاخص در اثر تغییرات، بیش‌تر از همه تغییرپذیر است؟ توضیح دهید.

۲ ــ در یک امتحان ریاضی ۱۵ نفر نمرات زیر را کسب نمودند :

۰،۱،۳،۱۴،۱۴،۱۵،۱۶،۱۶،۱۷،۱۷،۱۸،۱۸،۱۸،۱۹،۲۰

ــ میانگین، میانه و مد را به‌دست آورید.

ــ کدام شاخص نمایان‌گر بهتری از وضع دانش‌آموزان این کلاس می‌باشد؟

۳ــ میزان پرداخت حقوق در یک شرکت خصوصی در جدول زیر آمده است :

سِمَت	حقوق هر نفر (تومان)
مدیر شرکت	۳۰۰،۰۰۰
معاون	۲۰۰،۰۰۰
حسابدار	۸۵،۰۰۰
بازاریاب	۶۰،۰۰۰
منشی	۴۰،۰۰۰
۲ نفر فروشنده	۶۰،۰۰۰
مستخدم	۳۵،۰۰۰

مدیر شرکت در آگهی دعوت به همکاری در روزنامه اعلام می‌دارد که میانگین پرداخت حقوق در این شرکت بیش‌تر از ۱۰۰،۰۰۰ تومان است. آیا این رقم می‌تواند دلیلی برای پرداخت حقوق بالا در این شرکت باشد؟ چرا؟ کدام‌یک از شاخص‌ها برای نشان دادن وضع پرداختی کارکنان این شرکت بهتر می‌باشد چرا؟

۴ــ میانگین ۱۰ عدد ۳، ۸، ۵، ۹، ۳، ۷، ۸ ، a ، ۳، ۴ برابر ۵/۵ است.

الف ــ a را پیدا کنید.

ب ــ میانه را به‌دست آورید.

۵ــ مجموع هفت عدد متوالی برابر ۱۴۷ است. اگر میانگین این اعداد از میانه کم شود چه اتفاقی می‌افتد؟ (بدون محاسبه)

۶ــ نمرات درس فیزیک دانش‌آموزی در طول سال برابر است با :

۱۹، ۱۸/۵، ۷، ۱۸، ۱۹، ۱۷

الف ــ میانه و میانگین، مد را برای نمرات این دانش‌آموز حساب کنید.

ب ــ کدام‌یک از شاخص‌های فوق، بیان‌گر بهتری از وضع این دانش‌آموز در درس فیزیک است؟

ج ــ اگر معلم درس فیزیکِ این دانش‌آموز، برای جبران نمرهٔ ۷، امکان امتحان مجدد را به او بدهد، برای این‌که میانگین وی در این درس ۱۸/۵ شود، او در این امتحان چه نمره‌ای باید کسب کند؟

۷ــ اگر در مجموعه اعداد {x,۶۰,۸۰,۱۱۰,۷۰} میانگین برابر میانه و مد باشد، x را پیدا کنید.

۸ــ یک کارخانه تولید مواد غذایی می‌خواهد محصولی را بسته‌بندی کند. در یک تحقیق، مسئولین متوجه شدند که از ۲۲ نفر، پنج نفر ترجیح می‌دهند محصول در بسته‌بندی یکی‌ای ارائه شود، پنج نفر ترجیح دادند محصول در بسته‌بندی دوتایی و سه نفر در بسته‌بندی سه‌تایی و شش نفر در بسته‌بندی چهارتایی و سه نفر در بسته‌بندی شش‌تایی ارائه شود.

ــ میانگین، مد، میانه را برای تعداد محصول در هر بسته به‌دست آورید.

ــ به‌نظر شما این کارخانه باید از کدام شاخص در تصمیم‌گیری خود استفاده کند؟ چرا؟

۹ــ در بسیاری از دانشگاه‌ها، نمرات برحسب F, D, C, B, A اعلام می‌شوند. در این سیستم نمره‌گذاری A دارای وزن یا ارزش ۴، B دارای وزن یا ارزش ۳ و C دارای وزن یا ارزش ۲ و D دارای وزن یا ارزش ۱ و F وزن یا ارزش صفر دارد.

جدول زیر دروس و تعداد واحدها و نمرات اکتساب شده توسط مریم را در دانشگاه نشان می‌دهد :

ــ معدل مریم را در این ترم حساب کنید.

درس	تعداد واحد	نمره
ریاضیات	۳	B
فیزیک	۳	A
ورزش	۲	C
تاریخ	۲	B
ادبیات	۳	C
آزمایشگاه	۱	A

۱۰ــ یک باغبان می‌خواهد تعداد علف‌های هرز یک زمین بازی را تخمین بزند. او به‌طور تصادفی ۱۰۰ ناحیه از این زمین را هر یک به مساحت $۱۰۰cm^۲$ انتخاب نمود و تعداد علف‌های هرز هر ناحیه را شمرد. جدول زیر نتیجهٔ این شمارش است.

فراوانی نواحی	تعداد علف‌های هرز
۱۸	۰ــ۴
۲۵	۵ــ۹
۳۲	۱۰ــ۱۴
۱۴	۱۵ــ۱۹
۷	۲۰ــ۲۴
۴	۲۵ــ۲۹

ــ فراوانی تجمعی نواحی را به‌دست آورید.

ــ نمودار میله‌ای را برای فراوانی تجمعی نواحی رسم کنید.

ــ میانگین تعداد علف‌های هرز را برای هر قسمت (هر ۱۰۰ سانتی‌متر مربع) تخمین بزنید.

ــ اگر مساحت زمین $۸۰۰۰m^۲$ باشد، تعداد کل علف‌های هرز این زمین را تخمین بزنید.

۱۱ــ سن بازیکنان تیم ملی فوتبال یک کشور به شرح زیر است :

۲۲، ۲۳، ۱۸، ۳۱، ۱۹، ۲۹، ۲۶، ۲۶، ۲۴، ۲۷

۳۳، ۲۵، ۲۵، ۲۹، ۲۳، ۲۷، ۲۶، ۲۵

۲۵، ۲۶، ۲۱، ۳۱

ــ میانگین، مد و میانه سن بازیکنان این تیم را به‌دست آورید.

ــ تعداد بازیکنانی که سن آن‌ها بالاتر از میانگین است، بیش‌تر است یا تعداد بازیکنانی که سن آن‌ها از میانگین کم‌تر است؟

ــ تعداد بازیکنانی که سن آن‌ها بالاتر از میانه است بیش‌تر است یا تعداد بازیکنانی که سن آن‌ها از میانه کم‌تر است؟

ــ میانه و میانگین را در این بررسی مقایسه کنید. چه نتیجه‌ای می‌گیرید؟

۱۲ــ یک کارخانه سازنده اتومبیل‌های شخصی در نظر دارد اتومبیل‌های جدیدی را طراحی کند. برای این منظور اتومبیل‌هایی را که از یک چهارراه عبور می‌کردند، مورد بررسی قرار داد. اطلاعات به‌دست آمده به شرح زیر است :

۱۲ اتومبیل با ۱ سرنشین

۶ اتومبیل با ۲ سرنشین

۴ اتومبیل با ۳ سرنشین

۲ اتومبیل با ۴ سرنشین

۱ اتومبیل با ۶ سرنشین

ــ به‌طور متوسط هر اتومبیل چند سرنشین دارد؟

در این بررسی مد چند است؟

۱۳ــ داده‌های زیر مصرف برق ۵۰ خانوار ساکن تهران را نشان می‌دهد. (برحسب کیلووات ساعت)

۳۰، ۱۵، ۱۹، ۱۷، ۲۵، ۱۲، ۱۰، ۱۵، ۵، ۲۵

۷۵، ۶۸، ۸۸، ۹۸، ۶۸، ۶۶، ۵۵، ۴۷، ۳۵، ۲۵

۱۱۵، ۷۵، ۷۷، ۷۸، ۹۸، ۱۲۰، ۱۱۰، ۱۰۱، ۸۵، ۹۵

۱۵۰، ۱۳۵، ۱۴۰، ۱۲۰، ۸۸، ۹۳، ۷۵، ۶۸، ۴۵، ۶۶

۱۳۵، ۱۹۰، ۱۸۰، ۱۷۷، ۲۰۰، ۱۹۵، ۱۶۵، ۱۷۵، ۱۸۰، ۱۶۰

الف ــ آیا نمودار ساقه و برگ برای این داده‌ها مناسب است؟

ب ــ آیا میانگین معیار خوبی برای نشان دادن مصرف خانوارهای تهرانی هست یا خیر؟ توضیح دهید.

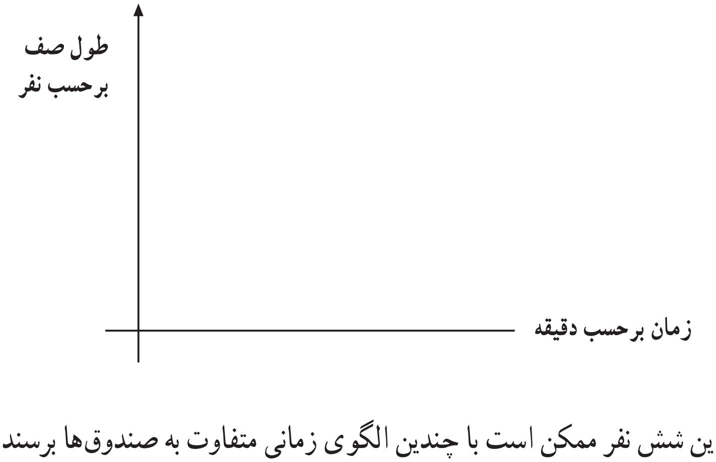
ج ــ نمودار جعبه‌ای این داده‌ها را رسم کنید.

د ــ تراکم مصرف بیش‌تر روی چه میزان از مصرف متمرکز شده است؟

۱۴ــ مدیر یک فروشگاه در نظر دارد برای حسابداران پای صندوق فروشگاهش برنامه‌ریزی کند. او می‌خواهد بداند در چه ساعاتی چند حسابدار برای رسیدگی به مشتریانش لازم دارد. برای این منظور او باید بداند به‌طور متوسط طول صف در ساعات مختلف روز چه‌قدر است. فرض کنید در ساعت ۱۰ صبح به‌طور متوسط هر ۱۰ دقیقه ۶ مشتری به صندوق مراجعه می‌کنند و فرض کنید رسیدگی به هر یک از آن‌ها به‌طور متوسط ۲ دقیقه طول می‌کشد. مثلاً اگر ترتیب رسیدن مشتری‌های A , B , C , D , E , F مطابق جدول زیر باشد :

زمان	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
مشتری	A	B	C,D			E				F

برای پیدا کردن میانگین طول صف در هر دقیقه باید نمودار طول صف بر حسب نفر در هر دقیقه را رسم کنیم.



امّا این شش نفر ممکن است با چندین الگوی زمانی متفاوت به صندوق‌ها برسند و رسم جدول بالا برای تمام الگوهای ممکن کاری بسیار دشوار و وقت‌گیر است. با توجه به الگوی بالا شما چه پیشنهادی برای کمک به مدیر فروشگاه می‌توانید بکنید؟

از ماشین حساب چگونه استفاده کنیم

پیدا کردن میانگین با استفاده از ماشین‌حساب: در یک بررسی از رانندگان در مورد تعداد تصادفاتی که در سال گذشته داشتند سؤال شد. اطلاعات به‌دست آمده در جدول زیر آمده است :

تعداد تصادفات	۰	۱	۲	۳	۴	۵	۶
تعداد رانندگان	۱۷	۱۳	۲۱	۴	۲	۲	۱

ــ چند راننده در این بررسی شرکت داشتند؟

۰ [Min]

۱۷ [+] ۱۳ = [M+]

۲۱ [M+]

۴ [M+]

۲ [M+]

۲ [M+]

۱ [M+] ⟶ [MR]

میانگین تعداد تصادفات برای هر راننده را به‌دست آورید :

۰ [Min]

۰ [×] ۱۷ = [M+]

۱ [×] ۱۳ = [M+]

۲ [×] ۲۱ = [M+]

۳ [×] ۴ = [M+]

۵ [×] ۲ = [M+]

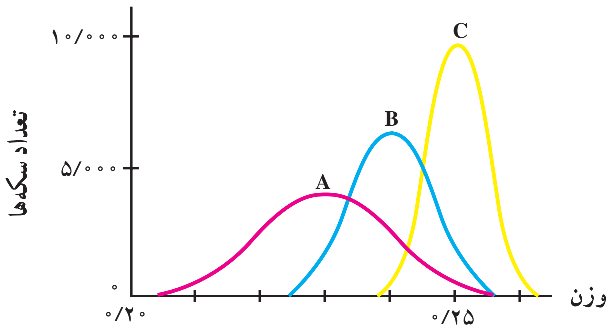
۶ [×] ۱ = [M+] ⟶ [MR] [÷] ۶۰ ≈ ۱/۳۸

از فرشتگانی که در آسمانهایت جایشان دادی، و از زمینِ خود برترشان بردی. آنان از دیگر آفریدگانت تو را بهتَر شناسند، و از عقاب تو بیشتر می‌هراسند، و به تو نزدیکترَند و نه پراکَندهٔ گردش زمان. آنان با مرتبتی که از آن برخوردارند، و از منزلتی که نزدِ تو دارند، و یکدله تو را دوست دارند و تو را فراوَان طاعت می‌گزارند، و اندک غفلتی در فرمان تو نیارند.



شاخص‌های پراکندگی

ضرابخانهٔ سکه‌های دولتی تصمیم گرفته است که وزن پنج تومانی‌هایی که در جریان بازار قرار می‌دهد، روی نمودار بیاورد. برای این منظور، چندین سکه پنج‌تومانی که تازه ضرب شده‌اند به‌طور تصادفی انتخاب و به دقت وزن می‌شوند. همچنین به همان تعداد سکه پنج تومـانی که پنج سال پیش به بازار عرضه شده بود و به‌همان تعداد سکهٔ پنج تومانی که ۱۰ سال پیش به بازار عـرضه شده بود به دقت وزن می‌شوند. سه نمـودار A وB و C بـه‌دست می‌آید.



ــ فکر می‌کنید کدام نمودار مربوط به سکه‌های نو ارائه شده به بازار و کدام نمودار مربوط به سکه‌های پنج ساله و ده ساله است؟

ــ با گذشت زمان میانگین وزن سکه‌های پنج تومانی چه تغییری می‌کند؟

ــ با گذشت زمان مد وزن سکه‌های پنج تومانی چه تغییری می‌کند؟

ــ چاق و یا لاغر شدن نمودارها را چگونه تعبیر می‌کنید؟



پراکندگی

در فصل قبل با شاخص‌های مرکزی آشنا شدیم. مثلاً شما براساس میانگین زمان‌هایی که طول کشیده است تا در ۲۰ روز گذشته از خانه به مدرسه بروید، از منزل خارج شده‌اید. در بین راه اعدادی را که معرف این مدت زمان بوده‌اند در ذهن خود مرور می‌کنید. میانگین اعداد ۳۵ دقیقه شده است و شما درست ۳۵ دقیقه قبل از شروع کلاس‌ها از خانه خارج شده‌اید، ولی اعدادی مانند ۴۰، ۳۸، ۳۰، ۳۲ و یا حتی ۴۵ در بین اعداد دیده می‌شود. شما مایل نیستید که دیر به مدرسه برسید، به ویژه آن‌که روز امتحان هم باشد. یادآوری اعداد بالا آسودگی خیال را به هم می‌ریزد. از خود می‌پرسید نکند امروز یکی از آن روزهایی باشد که بیش از ۳۵ دقیقه طول بکشد. اگر کمی بیش‌تر طول بکشد، (در حد ۵ دقیقه) اشکال عمده‌ای نخواهد بود ولی تأخیر یک ربع و یا بیش‌تر مشکل‌ساز است. هرچه‌قدر در بین داده‌های شما، اعداد دورتر از میانگین وجود داشته باشد، ناراحتی شما بیش‌تر خواهد بود. اگر قبلاً این اعداد را بهتر وارسی می‌کردید و صرفاً میانگین را مدنظر قرار نمی‌دادید الآن وضع بهتری داشتید، اگر از ابتدا درمی‌یافتید که اعداد تفاوت دارند می‌توانستید در این روز سرنوشت‌ساز تصمیم بهتری بگیرید. معیاری که می‌تواند این تفاوت‌ها و میزان آن‌ها و به‌خصوص دوری آن‌ها از میانگین را برای ما اندازه‌گیری کند شاخص پراکندگی نامیده می‌شود. ما به چند شاخص پراکندگی در زیر اشاره می‌کنیم :

دامنهٔ تغییرات

دامنهٔ تغییرات که در فصل دوم با آن آشنا شدید یکی از شاخص‌های پراکندگی است. هرچه قدر داده‌ها از هم دورتر باشند، دامنهٔ تغییرات آن‌ها بیش‌تر خواهد بود. دیدیم اگر دامنهٔ تغییرات برابر صفر باشد تمام داده‌ها با هم برابرند. یعنی هیچ تفاوتی بین آن‌ها موجود نیست. حال به دو دسته دادهٔ زیر که نمره‌های دانش‌آموزان در درس فیزیک در دو نمونهٔ ۱۰ عضوی را به نمایش می‌گذارند توجه کنید.

الف) ۱۹، ۱۸، ۱۷، ۱۶، ۱۵، ۱۴، ۱۳، ۱۲، ۱۱، ۱۰

ب) ۱۹، ۱۴، ۱۵، ۱۵، ۱۵، ۱۵، ۱۴، ۱۴، ۱۵، ۱۰

دامنهٔ تغییرات هر دو دسته یک اندازه است :

$$R = ۱۹ - ۱۰ = ۹$$

ولی پراکندگی داده‌های دستهٔ دوم کم‌تر است. در داده‌های دسته دوم ما عملاً اعداد ۱۴ و ۱۵ را داریم. به ۱۰ و ۱۹ بیش‌تر باید با شک و تردید نگریست. این دو مقدار ممکن است به خاطر اندازه‌گیری‌های نادرست حاصل شده باشند و یا ممکن است علل دیگری برای وجود آن‌ها در بین داده‌ها باشد. اما در داده‌های دستهٔ اول این دو مقدار کاملاً طبیعی به نظر می‌رسند، زیرا مقادیر نزدیک به آن‌ها نیز داریم. با این توضیحات ما در داده‌های دسته دوم بیش‌تر روی ۱۴ و ۱۵ حساب می‌کنیم و این دو مقدار هم به هم خیلی نزدیک‌اند. این توضیح می‌رساند که دامنهٔ تغییرات در بعضی مواقع به خوبی نمی‌تواند پراکندگی موجود در داده‌ها را نشان دهد. بنابراین به دنبال معیاری هستیم که بهتر از دامنه تغییرات بتواند پراکندگی و تفاوت‌ها را نشان دهد. قبل از آن که به معرفی معیار بعدی برویم ذکر یک نکته ضروری است که وقتی با دید مجرد ریاضی به دو دسته عدد بالا نگاه می‌کنید تفاوت هر دو را در حد ۹ می‌بینید، ولی با دید آماری مسئله کمی فرق می‌کند. در آمار حضور دو مقدار ۱۰ و ۱۹ را در دستهٔ «ب» زیاد جدی نمی‌گیریم چون این‌ها مقادیر تک افتاده‌ای هستند. به‌نظر می‌رسد این دو مقدار به جامعه تعلق نداشته باشند. ما در این درس سعی می‌کنیم علاوه‌بر نقطه نظراتی که شما دربارهٔ مسائل اطراف خود دارید پایه‌های دید آماری را نیز ایجاد کنیم و آن را توسعه بخشیم.





چارک‌های بالا و پایین

گفتیم که دامنهٔ تغییرات ممکن است در بعضی مواقع تعبیرهای نامناسب از جامعه ارائه کند، مثلاً در جامعه‌ای که عملاً داده‌ها به هم نزدیک‌اند، به‌علت وجود دو مقدار خیلی کوچک و خیلی بزرگ در جامعه، دامنهٔ تغییرات عدد بزرگی به‌دست آید، و حال آن‌که جامعه از دید آماری جامعه متمرکزی باشد. تعداد بازدیدکنندگان از کتابخانهٔ مدرسه در ۱۲ روز کاری در زیر آمده است.

۱۱، ۱۵، ۶، ۱۰، ۹، ۵، ۶، ۷، ۸، ۲، ۱، ۰

دامنهٔ تغییرات این داده‌ها برابر ۱۵ است و به نظر زیاد می‌رسد ولی داده‌ها عموماً از ۵ تا ۱۱ پراکنده شده‌اند که پراکندگی آن‌ها زیاد نیست. از این رو برای حذف تأثیر داده‌های بزرگ و کوچک روی تعدادی از اعداد بالای داده‌ها و تعدادی از اعداد پایین داده‌ها را حذف می‌کنند. این‌که چه تعدادی حذف شود بستگی به تعداد داده‌ها و نزدیکی داده‌ها به‌هم دارد. در بعضی مواقع یک دهم داده‌ها از بالا و یک دهم داده‌ها از پایین را حذف می‌کنند و در برخی مواقع یک چهارم داده‌های بالایی و یک‌چهارم داده‌های پایینی را حذف می‌کنند و بعد از آن دامنه تغییرات داده‌های باقی‌مانده را محاسبه می‌کنند. به‌عنوان مثال در داده‌های بالا اگر یک چهارم داده‌ها را از بالا و پایین حذف کنیم به داده‌های زیر خواهیم رسید :

۱۵، ۱۱، ۱۰، **۹، ۸، ۷، ۶، ۶، ۵**، ۲، ۱، ۰

بنابراین دامنهٔ تغییرات این داده‌ها عبارت خواهد بود از $۹ - ۵ = ۴$ و حال آن‌که دامنه تغییرات داده‌های اصلی برای $۱۵ - ۰ = ۱۵$ است. اگر توجه کرده باشید در برخی از مسابقات امتیازهایی را که داورها می‌دهند، بیش‌ترین و کم‌ترین امتیاز را حذف می‌کنند و بر اساس امتیازهای باقی‌مانده امتیاز ورزشکار را تعیین می‌کنند.

تمرین

۱ــ دو کارخانه تولید کننده مواد غذایی A و B شکلات در بسته‌بندی ۴۸ گرمی می‌فروشند. پنج بسته شکلات به‌صورت تصادفی از یک فروشگاه مواد غذایی از دو محصول انتخاب شد و تمام وزن بسته‌ها به دقت اندازه‌گیری شد.

نتیجه زیر به دست آمد :

A:۴۸/۰۸ ،۴۸/۳۲ ،۴۷/۹۶ ،۴۷/۸۴ ،۴۷/۹۶

B:۴۹/۱۶ ،۴۸/۸۴ ،۴۸/۸۸ ،۴۹/۰۸ ،۴۹

الف ــ کدام کارخانه شکلات بیش‌تری در بسته‌ها می‌فروشد؟ از چه شاخصی استفاده می‌کنید؟

ب ــ کدام کارخانه در توزیع شکلات یکنواخت‌تر عمل کرده است؟

۲ــ اگر دامنهٔ تغییرات برابر صفر باشد دربارهٔ داده‌ها چه نتیجه‌ای می‌گیرید؟

۳ــ تغییر واحد اندازه‌گیری چه تأثیری در دامنه تغییرات می‌گذارد؟ مثلاً اگر شما داده‌ها را برحسب متر اندازه‌گیری کرده باشید و سپس آن‌ها را برحسب سانتی‌متر به‌کار برید دامنه تغییرات چگونه تغییر می‌کند؟

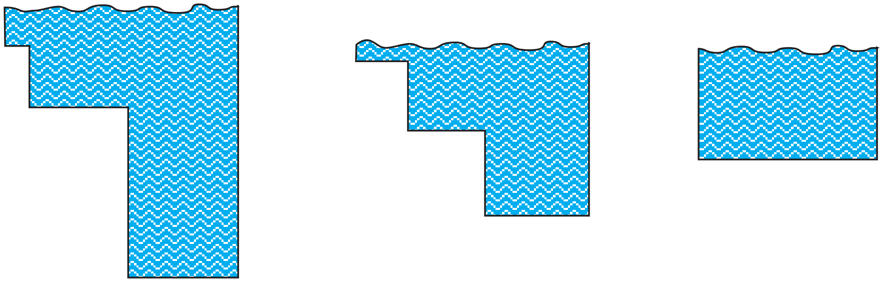
۴ــ اگر به داده‌ها یک مقدار ثابت اضافه کنیم دامنهٔ تغییرات چگونه تغییر می‌کند؟



واریانس

اگر شما خوب شنا نمی‌دانستید و می‌خواستید در استخری شنا کنید که عمق آن متغیر است برای این‌که اطمینان حاصل کنید که در خطر نخواهید بود چه اطلاعاتی را لازم دارید؟

اگر میانگین عمق استخر را به شما داده باشند و این میانگین برابر ۱/۵ متر باشد، آیا برای شنا در این استخر اطمینان دارید؟ به حالات زیر توجه کنید.



در هر سه استخر بالا میانگین عمق برابر ۱/۵ متر است. می‌خواهیم با یک شاخص عددی پراکندگی متغیر عمق را نسبت به عمق میانگین بررسی کنیم. اگر بدانیم که تغییرات عمق در اطراف عمق میانگین، یعنی ۱/۵ متر زیاد نیست با اطمینان بیش‌تری در آن شنا خواهیم کرد. به شرط آن که این پراکندگی با یک شاخص عددی مناسب معرفی شده باشد.

دامنه تغییرات شاخص مناسبی است. اما در تصمیم گیری‌های کلان از ارزش آماری زیادی برخوردار نیست. زیرا ما به شاخص‌هایی نیاز داریم که هم پراکندگی داده‌ها و هم فراوانی آن‌ها را مد نظر قرار دهند.

دیدیم که پراکندگی یعنی این که داده‌ها از مرکز خود چه‌قدر دور هستند، پس یک راه ابتدایی ممکن است این‌طور به نظر برسد که تک‌تک مقادیر را از میانگین کم کنیم. این تفاضل‌ها را *انحراف از میانگین* می‌نامیم و سپس مجموع مقادیر حاصل را به‌دست آوریم، یعنی مجموع مقادیر

$$x_1 - \bar{x} \text{ و } x_2 - \bar{x} \text{ و } \dots \text{ و } x_n - \bar{x}$$

را حساب کنیم. قبلاً دیده‌ایم که مجموع مقادیر بالا برابر صفر است. لذا برای هر نوع داده‌ای اعم از آن که داده‌ها به میانگین نزدیک باشند و یا از آن دور باشند، آن چه که از این روش به‌دست می‌آید برابر صفر است. پس معیار مذکور نمی‌تواند معیار مفیدی باشد. البته علت آن که این مجموع برابر صفر شد به این خاطر است که برخی از داده‌ها از میانگین بزرگ‌تر و برخی دیگر کوچک‌ترند. درنتیجه مقادیر مثبت و منفی حاصل می‌شوند که مجموع آن‌ها یک‌دیگر را خنثی می‌کنند. برای جلوگیری از علامت‌های مثبت و منفی یکی از راه‌ها استفاده از قدرمطلق است. یعنی از مقادیر $|x_n - \bar{x}|$ و ... و $|x_1 - \bar{x}|$

استفاده کنیم. این مقادیر می‌تواند معیاری برای پراکندگی ارائه کند ولی معمولاً کارکردن با قدرمطلق کار آسانی نیست و از این رو از توان دوم مقادیر بالا استفاده می‌کنیم یعنی مجموع مجذورات فاصله از میانگین را در نظر می‌گیریم. پس یک معیار پراکندگی ممکن است به‌صورت زیر باشد :

$$(x_1-\bar{x})^2+\cdots+(x_n-\bar{x})^2$$

ملاحظه می‌شود که اگر داده‌ها برابر باشند مقدار بالا برابر صفر است و بالعکس اگر مقدار بالا برابر صفر باشد داده‌ها برابر بوده و در نتیجه هیچ تفاوتی در بین آن‌ها نیست. اما اشکالی که دستور بالا دارد آن است که اگر تعدادی دادهٔ دیگر اضافه کنیم به مجموع بالا مقادیر مثبت دیگری اضافه خواهد شد و لذا این مجموع بزرگ‌تر خواهد شد. حال آیا درست است که بگوییم با اضافه شدن داده‌ها پراکندگی بیش‌تر می‌شود؟ مسلماً پاسخ منفی است، چون پراکندگی چیزی است که به اصل جامعه مربوط است و این نوع محاسبات نباید در مقدار آن‌ها تأثیر بگذارد. پس برای آن که تأثیر تعداد داده‌ها را از بین ببریم مجموع مذکور را بر تعداد آن‌ها تقسیم می‌کنیم و از این‌جا به معیار زیر می‌رسیم :

واریانس

واریانس برابر میانگین مجذور انحرافات از میانگین است و آن را با σ^2 نشان می‌دهیم. بنابراین :

$$\sigma^2=\frac{(x_1-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n}$$

(σ از حروف کوچک یونانی است و سیگما خوانده می‌شود. حرف بزرگ آن Σ است)



تعداد ساعاتی که ۴ دانش‌آموز در طول هفته به ورزش اختصاص داده‌اند در زیر آمده است : واریانس این داده‌ها را حساب کنید.

۱ ,۵ ,۷ ,۹

ابتدا میانگین آن‌ها را حساب می‌کنیم.

$$\text{میانگین} : x = \frac{۱+۵+۷+۹}{۴} = \frac{۲۲}{۴} = ۵/۵$$

سپس انحراف از میانگین را حساب می‌کنیم.

انحرافات از میانگین : $-۴/۵,\ -۰/۵,\ ۱/۵,\ ۳/۵$

مجذور انحرافات از میانگین : $۲۰/۲۵,\ ۰/۲۵,\ ۲/۲۵,\ ۱۲/۲۵$

$$\text{میانگین مجذور انحرافات از میانگین} : \frac{۲۰/۲۵+۰/۲۵+۲/۲۵+۱۲/۲۵}{۴} = \frac{۳۵}{۴} = ۸/۷۵$$

پس $\sigma^۲ = ۸/۷۵$

اگر داده‌ها برابر باشند واریانس آن‌ها صفر است و بالعکس.

اگر داده‌ها برابر باشند، میانگین آن‌ها نیز همان مقدار مشترک داده‌ها خواهد بود. پس انحرافات از میانگین برابر صفر است. لذا مجموع مجذورات انحراف صفر خواهد بود. پس واریانسی که از تقسیم این مجموع بر تعداد داده‌ها به‌دست می‌آید برابر صفر است. حال فرض کنید واریانس صفر باشد. پس باید مجموع مجذورات انحرافات صفر شده باشد. مجذور انحرافات عددی نامنفی است و چون مجموع آن‌ها صفر است لازم می‌آید هریک از آن‌ها صفر باشند، چون در غیر این صورت مجموع مجذور انحرافات صفر نخواهد شد. با صفر شدن هریک از انحرافات نتیجه می‌شود که تمام داده‌ها برابر میانگین هستند. پس مساوی‌اند.



نکاتی دربارهٔ واحد میانگین و واریانس : فرض کنید مطالعه‌ای دربارهٔ قد افراد انجام می‌دهید، و داده‌های شما برحسب سانتی‌متر اندازه‌گیری و بیان شدند. از آن‌جایی که میانگین به‌وسیله مجموع داده‌ها به‌دست می‌آید آن هم برحسب سانتی‌متر خواهد بود. یعنی واحد میانگین از نوع همان واحد داده‌هاست. اما در واریانس شرایط فرق می‌کند زیرا در این‌جا شما مقادیر $(x_i - \bar{x})$ را که برحسب سانتی‌متر است به توان ۲ می‌رسانید، پس مجذور انحرافات بر حسب سانتی‌مترمربع است و لذا مجموع آن‌ها و درنتیجه واریانس برحسب سانتی‌مترمربع خواهد بود. بنابراین واحد واریانس از نوع مجذور واحد متغیر است.

دستور دیگری برای محاسبه واریانس: دیدیم که برای محاسبهٔ واریانس باید میانگین مجذور انحرافات از میانگین را حساب کنیم، یعنی

$$\sigma^2 = \frac{1}{n}\left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2\right]$$

محاسبه واریانس با استفاده از این دستور در بعضی مواقع مشکلاتی دارد.
می‌توان ثابت کرد عبارت بالا را می‌توان به‌صورت زیر نیز نوشت

$$\sigma^2 = \frac{1}{n}\left[x_1^2 + \cdots + x_n^2\right] - \bar{x}$$

اثبات این دستور چندان مشکل نیست، توجه کنید که

$$\sigma^۲ = \frac{۱}{n}\left[(x_۱-\bar{x})^۲+\cdots+(x_n-\bar{x})^۲\right]$$

$$=\frac{۱}{n}\left[(x_۱^۲-۲x_۱\bar{x}+\bar{x}^۲)+\cdots+(x_n^۲-۲x_n\bar{x}+\bar{x}^۲)\right]$$

$$=\frac{۱}{n}\left[(x_۱^۲+\cdots+x_n^۲)-۲(x_۱+\cdots+x_n)\bar{x}+n\bar{x}^۲\right]$$

$$=\frac{x_۱^۲+\cdots+x_n^۲}{n}-\frac{۲(x_۱+\cdots+x_n)}{n}\bar{x}+\bar{x}^۲$$

$$=\frac{x_۱^۲+\cdots+x_n^۲}{n}-۲\bar{x}^۲+\bar{x}^۲$$

$$=\frac{x_۱^۲+\cdots+x_n^۲}{n}-\bar{x}^۲$$

توجه کنید در $x_۱^۲+\cdots+x_n^۲$ ابتدا مقادیر x_i ها را به توان ۲ می‌رسانیم و سپس آن‌ها را با هم جمع می‌کنیم و در $\bar{x}^۲$ ، $\bar{x}$ (میانگین داده‌ها) را به توان دو می‌رسانیم.

حال واریانس داده‌های زیر را حساب می‌کنیم.

۱ ،۵ ،۶ ،۷ ،۹

ابتدا میانگین را حساب می‌کنیم :

$$\bar{x}=\frac{۱+۵+۶+۷+۹}{۵}=\frac{۲۸}{۵}=۵/۶$$

$$x_۱^۲+\ldots+x_۵^۲=۱+۲۵+۳۶+۴۹+۸۱$$

$$=۱۹۲$$

پس :

$$\sigma^۲=\frac{۱۹۲}{۵}-(۵/۶)^۲$$

$$=۳۸/۴-۳۱/۳۶=۷/۰۴$$

با استفادهٔ مستقیم از دستور واریانس نیز به‌همین نتیجه می‌رسیدیم

$$\sigma^۲=\frac{(۱-۵/۶)^۲+(۵-۵/۶)^۲+(۶-۵/۶)^۲+(۷-۵/۶)^۲+(۹-۵/۶)^۲}{۵}$$

$$=\frac{۲۱/۱۶+۰/۳۶+۰/۱۶+۱/۹۶+۱۱/۵۶}{۵}$$

$$=۷/۰۴$$



انحراف معیار

این که واحد واریانس از نوع مجذور واحد متغیر است می‌تواند مشکلات و سوء تفاهماتی را در پی داشته باشد. فرض کنید شما قدها را برحسب متر اندازه‌گیری کرده باشید و همکلاسی شما قدها را برحسب سانتی‌متر، در این صورت واریانسی که همکلاسی شما حساب می‌کند ۱۰۰۰۰ برابر واریانسی است که شما حساب کرده‌اید. همکلاسی شما ممکن است این‌طور نتیجه بگیرد که پراکندگی در قد بسیار زیاد و چشم‌گیر است و حال آن که نظر شما ممکن است غیر از این باشد. برای آن که این اختلاف نظرها را از بین ببریم سعی می‌کنیم که تفاوت عمده در واحد واریانس و واحد میانگین را با جذر گرفتن از واریانس از بین ببریم. جذر واریانس را انحراف معیار می‌گوییم.

انحراف معیار که با نماد σ نشان داده می‌شود برابر جذر واریانس است.

واحد انحراف معیار همان واحد متغیر است.

تعداد ساعت‌هایی در هفته که ۴ نفر به‌صورت داوطلبانه دریک بیمارستان خدمت می‌کنند در زیر آمده است. انحراف معیار داده‌ها را حساب کنید.   ۱ ،۳ ،۴ ،۵ ،۶

$$\bar{x} = \frac{۱+۳+۴+۵+۶}{۵} = \frac{۱۹}{۵} = ۳/۸$$

$$\sigma^۲ = \frac{(-۲/۸)^۲ + (-۰/۸)^۲ + (۰/۲)^۲ + (۱/۲)^۲ + (۲/۲)^۲}{۵}$$

$$= \frac{۷/۸۴ + ۰/۶۴ + ۰/۰۴ + ۱/۴۴ + ۴/۸۴}{۵}$$

$$= \frac{۱۴/۸۰}{۵} = ۲/۹۶$$

در نتیجه: $\sigma = \sqrt{۳/۹۶} \approx ۱/۷$

تمرین

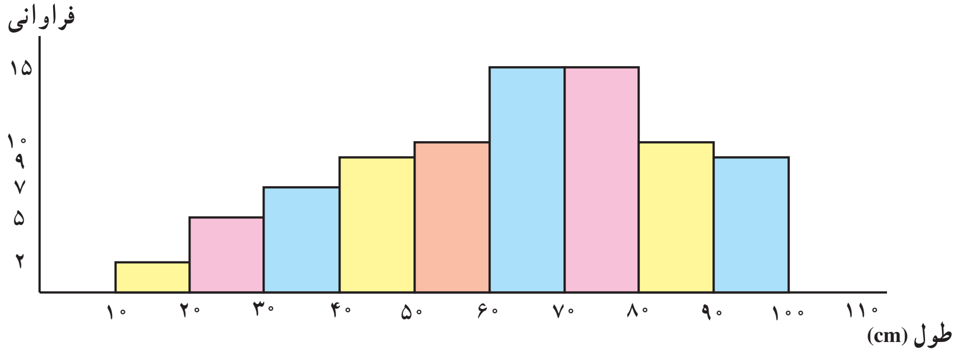
۱ــ نمودار زیر طول ۸۲ گیاه پرورش داده شده در یک آزمایشگاه را نشان می‌دهد.

ــ طول چند گیاه بین ۵۰ تا ۶۰ سانتی‌متر است؟

ــ طول چند گیاه بین ۷۰ تا ۹۰ سانتی‌متر است؟

ــ میانگین و انحراف معیار طول گیاهان را به‌دست آورید.

ــ توضیح دهید چه ویژگی نمودار بیان‌گر آن است که میانه و میانگین برابر نیستند.



۲ــ درجه حرارت بدن ۸ بیمار (به نزدیک‌ترین عدد صحیح گرد شده است) در زیر آمــده است :

۳۸ ۳۸ ۳۹ ۳۹ ۴۰ ۴۰ ۴۱ ۴۱

میانگین و انحراف معیار را حساب کنید.

درجه حرارت‌های واقعی به قرار زیر می‌باشند :

۳۸/۰ ۳۸/۴ ۳۹/۰ ۳۹/۴ ۳۹/۸ ۴۰/۲ ۴۰/۹ ۴۱/۲

ــ میانگین و انحراف معیار را حساب کنید.

ــ گرد کردن داده‌ها چه تأثیری در میانگین و انحراف معیار داشت؟

ــ آیا گرد کردن همواره تأثیری بر میانگین و انحراف معیار دارد؟

۳ــ در یک نمونهٔ ۹۰۰ نفری از مردان دربارهٔ مسافتی که روزانه از منزل به محل کار طی می‌کنند سؤال شد. اعداد صفحه بعد نتیجهٔ این بررسی را نشان می‌دهد.

مسافت طی شده (کیلومتر)	درصد
کمتر از ۱	۱۷
۱ و کمتر از ۳	۳۱
۳ و کمتر از ۸	۳۸
۸ و کمتر از ۱۰	۷
۱۰ و کمتر از ۱۵	۵
۱۵ و کمتر از ۵۰	۲
	۱۰۰

ــ نمودار مستطیلی این داده‌ها را رسم کنید.

ــ میانگین و انحراف معیار را برای این داده‌ها به‌دست آورید.

ــ اگر میانگین مسافت طی شده توسط زنان کارمند ۵ کیلومتر باشد، با فرض آن که انحراف معیار هر دو گروه یکسان است، آیا می‌توانید بگویید که مردان فاصلهٔ بیش‌تری طی می‌کنند یا زنان؟

۴ــ دو گروه داده به‌صورت زیر داریم :

گروه اول : ۶ ۵ ۴ ۳ ۲

گروه دوم : ۸ ۷ ۶ ۵ ۴

الف ــ واریانس هر گروه را حساب کنید.

ب ــ این دو گروه داده چه رابطه‌ای با هم دارند؟

ج ــ رابطه موجود بین دو گروه داده، چه اثری روی واریانس‌ها داشته است؟

د ــ با توجه به تعریف واریانس، نتیجهٔ حاصل از قسمت ج را توضیح دهید.

۵ــ دو گروه داده به‌صورت زیر داریم :

گروه اول : ۵ ۴ ۳ ۲ ۱

گروه دوم : ۱۵ ۱۲ ۹ ۶ ۳

الف ــ واریانس هر گروه را حساب کنید.

ب ــ این دو گروه چه رابطه‌ای با هم دارند؟

ج ــ رابطه موجود بین دو گروه داده، چه تأثیری روی واریانس داشته است؟

د ــ نتیجه حاصل از قسمت ج را توضیح دهید.

۶ــ در حالت کلی ثابت کنید که اگر داده‌ها را با عددی جمع کنیم، در واریانس آن‌ها تغییری حاصل نخواهد شد. به زبان نمادها $\sigma_{a+x}^{2} = \sigma_{x}^{2}$

۷ــ در حالت کلی ثابت کنید که اگر داده‌ها را در عددی ضرب کنیم واریانس آن‌ها در مجذور این عدد ضرب خواهد شد. به زبان نمادها $\sigma_{ax}^{2} = a^{2}\sigma_{x}^{2}$

۸ــ با ترکیب دو تمرین بالا نتیجه بگیرید که $\sigma_{ax+b}^{2} = a^{2}\sigma_{x}^{2}$

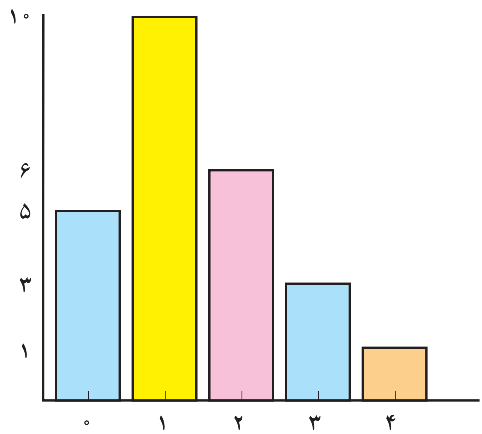
۹ــ کدامیک از برابری‌های زیر دربارهٔ انحراف معیار صحیح است؟

$\sigma_{ax+b} = a\sigma_{x} \quad \sigma_{ax+b} = -a\sigma_{x} \quad \sigma_{ax+b} = |a|\sigma_{x}$

۱۰ــ در یک بررسی تعداد فرزندان ۲۵ نفر از زنان مورد سؤال بود. نتیجه در نمودار زیر نشان داده شده است.

الف ــ نشان دهید متوسط تعداد فرزندان برابر ۱/۴ است.

ب ــ نشان دهید انحراف معیار این داده‌ها تقریباً ۱/۰۶ است.



ــ گروه ۲۵ نفری دیگری از زنان مورد بررسی قرار گرفتند. در این گروه متوسط تعداد فرزندان ۲/۴ و انحراف معیار آن‌ها ۲ است. با توجه به اطلاعات به‌دست آمده در الف و ب، تفاوت بین تعداد فرزندان این دو گروه را بررسی کنید.

ضریب تغییرات

مسلماً قد همکلاسی‌های شما متفاوت است، همچنین وزن آن‌ها نیز چنین است ولی در سن همکلاسی‌های شما اختلاف زیادی نیست. یعنی انتظار می‌رود واریانس سن کم، ولی واریانس قد و وزن زیاد باشد. حال سؤال این است که تغییرات در قد همکلاسی‌ها بیش‌تر است یا در وزن آن‌ها. سن تقریباً ثابت است، زیاد متغیر نیست ولی می‌خواهیم بدانیم از قد و وزن کدام بیش‌تر متغیر است. اگر این دو صفت برحسب یک واحد می‌بودند مقایسه واریانس آن‌ها می‌توانست جواب‌گوی این سؤال باشد ولی چون واحدها یکی نیستند مقایسه میزان پراکندگی این دو متغیر از طریق واریانس‌ها ممکن نیست. برای آن‌که بتوانیم به سؤال بالا جواب دهیم باید شاخصی معرفی کنیم که بدون واحد باشد تا امکان مقایسه فراهم آید. اگر به خاطر داشته باشید در جدول فراوانی برای آن که تأثیر اندازه نمونه را از بین ببریم به محاسبهٔ درصد فراوانی نسبی پرداختیم. در این‌جا نیز برای از بین بردن واحد اندازه‌گیری از معیار **ضریب تغییرات** استفاده می‌کنیم.

ضریب تغییرات که با نماد CV نشان می‌دهیم عبارت است از خارج قسمت انحراف معیار بر میانگین



تذکر: از آن‌جایی که ضریب تغییرات معیاری برای میزان پراکندگی است باید مثبت باشد، چون پراکندگی منفی معنی ندارد. پس لازم است $\bar{x}$ ، که در مخرج کسر بالا آمده است، مثبت باشد. برای اطمینان از مثبت بودن $\bar{x}$ ، ضریب اطمینان را فقط برای داده‌های مثبت تعریف می‌کنیم.

ضریب تغییرات داده‌های زیر را محاسبه می‌کنیم :

۱، ۳، ۴، ۵، ۶

قبلاً دیدیم که $\bar{x} = ۳/۸$ و $\sigma = ۱/۷$ پس

$$CV = \frac{۱/۷}{۳/۸}$$

تذکر: هرچه‌قدر داده‌ها بزرگ‌تر باشند زمانی پراکندگی در آن‌ها محسوس است که تفاوت داده‌ها بیش‌تر باشد. ما اختلاف بین ۳ و ۳/۵ را به‌خوبی درک می‌کنیم ولی اختلاف بین ۳۰۰۰ و ۳۰۰۰/۵ چندان برای ما معنی‌دار نیست. شما در خرید اجناس ارزان قیمت انتظار تخفیف زیاد ندارید ولی در خرید اجناس گران قیمت از بحث و مجادله بر روی مبالغ کم خودداری می‌کنید. دلیل این امر آن است که در مقادیر بزرگ می‌توان از اختلاف‌های کم صرف‌نظر کرد. مثلاً از نظر آماری داده‌های زیر تقریباً با هم برابرند و تفاوت زیادی با هم ندارند.

۱۱۱، ۱۱۱/۵، ۱۱۲، ۱۱۲/۵

ولی داده‌های زیر با هم تفاوت دارند

۱، ۱/۵، ۲، ۲/۵

وقتی از ۱ به ۱/۵ می‌رویم پنجاه درصد ۱ به آن اضافه می‌شود ولی وقتی که از ۱۱۱ به ۱۱۱/۵ می‌رویم حدود چهارهزارم ۱۱۱ به آن اضافه می‌شود که ناچیز است. بنابراین پراکندگی به طریقی باید به‌نسبت بزرگی داده‌ها تعدیل شود. از این رو اگر ما انحراف معیار را بر $\bar{x}$ تقسیم کنیم میزان پراکندگی را برای یک واحد از میانگین حساب کرده‌ایم و به این ترتیب تأثیر بزرگی داده‌ها را از بین برده‌ایم. با ملاحظه بالا می‌توان توصیف زیر را برای ضریب تغییرات به‌کار برد.

ضریب تغییرات عبارت است از میزان پراکندگی به ازای یک واحد از میانگین

تمرین

۱ــ اگر میانگین برابر ۴ و انحراف معیار برابر ۶ باشد ضریب تغییرات چه‌قدر است؟

۲ــ ضریب تغییرات سن دانش‌آموزان کلاس شما ۱۰ سال دیگر چه تغییری می‌کند؟ کم‌تر می‌شود یا بیش‌تر؟

۳ــ اگر ۲۰٪ نمره هر دانش‌آموز به نمره او اضافه شود چه تأثیری روی ضریب تغییرات نمره‌ها حاصل می‌شود؟

۴ــ دو دسته دادهٔ زیر قیمت کالایی را در دو بازار جداگانه نشان می‌دهد.

۱۰، ۱۱، ۱۰، ۹، ۱۲، ۱۱، ۱۰، ۱۲، ۹، ۱۳، ۸

۱۰، ۹، ۱۱، ۱۰، ۱۰، ۱۱، ۹، ۱۰، ۸، ۱۳، ۱۰

الف ــ در کدام دسته پراکندگی بیش‌تر مشهود است؟

ب ــ دامنهٔ تغییرات را در هر دو دسته حساب کنید.

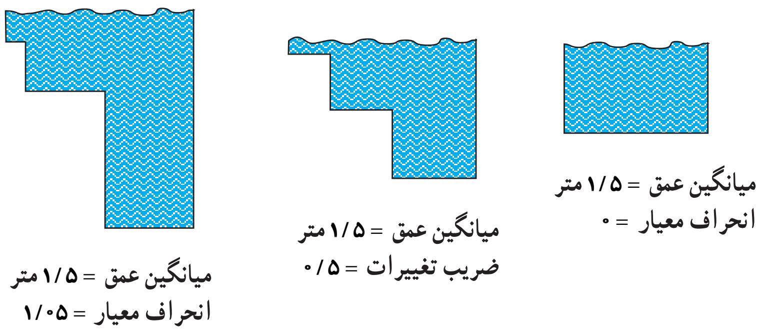
ج ــ آیا مقدار دامنهٔ تغییرات با پراکندگی مشاهده شده در «الف» هم‌خوانی دارد؟

د ــ ترجیح می‌دهید از کدام بازار خرید کنید؟ چرا؟

هــ اگر داده‌های واقعی در اختیار شما نباشند، فقط به صرف داشتن دامنهٔ تغییرات می‌توانستید تصمیم‌گیری کنید؟

و ــ کدام یک از شاخص‌های پراکندگی در تصمیم‌گیری شما می‌تواند مؤثرتر باشد؟ این شاخص را معرفی کرده و محاسبه کنید.

۵ــ در محله‌ای سه استخر شنا با مشخصات زیر موجود است :



الف ــ برای فردی مبتدی با قد ۱/۴۵ متر شنا کردن در کدام یک از این استخرها امنیت بیش‌تری دارد؟ چرا؟

ب ــ در کدام استخر با خطر بیش‌تری روبه‌رو است؟ چرا؟

۶ــ دو دسته دادهٔ زیر ارزش واردات برحسب کشورهای مبدأ را در سال‌های ۶۵ و ۷۰ نشان می‌دهد. (میلیون ریال)

کشور و منطقه	۱۳۵۵	۱۳۶۰	۱۳۶۵	۱۳۶۶	۱۳۶۷	۱۳۶۸	۱۳۶۹	۱۳۷۰
امارات	۳۲۴۳	۳۲۳۷۱	۴۱۳۴۳	۲۷۴۰۸	۱۹۱۹۹	۶۸۶۷۵	۶۵۴۵۲	۹۵۶۶۱
اندونزی	۴۱	۳۴	۲۱۱	۱۱۰۵	۷۳۴	۵۰۱۵	۴۳۳۵	۶۹۱۱
بنگلادش	۱۰۵۰	۳۰۹۸	۲۲۲۸	۲۶۷۷	۱۱۸۶	۳۱۹۶	۳۳۲۴	۳۰۹۶
پاکستان	۴۴۶۸	۶۷۲۱	۷۰۴۱	۴۳۰۴	۲۷۴۲	۱۰۴۵۵	۳۲۹۶	۱۲۴۰۳
تایلند	۱۳۷۵	۱۹۸۷۵	۴۰۲۵	۹۵۶۹	۴۸۹۳	۱۳۴۳۳	۱۳۲۷۷	۷۲۰۶
تایوان	۹۲۰	۱۰۵۱۰	۷۳۳۰	۴۶۱۹	۴۷۵۴	۵۲۹۱	۲۸۹۸	۱۰۲۹۸
ترکیه	۲۹۳۷	۲۳۳۶۹	۴۵۲۱۶	۳۳۳۳۱	۲۹۵۲۹	۵۰۶۴۵	۴۸۹۳۶	۶۹۶۳۳
چین	۷۰۹۳	۹۶۳۱	۱۹۹۳	۳۷۴۷	۲۴۸۶	۳۴۶۰	۱۱۶۱۳	۱۹۲۶۲
ژاپن	۱۵۵۵۰۰	۱۲۹۶۰۶	۹۷۶۲۷	۷۴۰۶۸	۵۸۰۶۶	۷۰۴۳۹	۱۳۰۲۴۸	۲۱۶۹۲۵
سنگاپور	۲۷۷۳	۸۲۵۴	۵۷۰۰	۳۸۱۶	۵۳۲۸	۵۵۵۴	۷۴۴۶	۹۰۳۰
سوریه	۲۷۳	۱۷۸	۹۱۴	۲۵۸۲	۶۸۸۳	۸۸۷۸	۲۰۰۵	۹۰۷
عربستان	۲۲	۱۶۹۱	۲۰۶۶	۷۳۲	۹۱۰	۶۳۸۲	۴۶۱	۲۰۸۶

ــ تغییرات ارزش واردات در کدام سال بیش‌تر بوده‌است؟

۷ــ با مراجعه به جدول صفحه بعد به سؤالات زیر پاسخ دهید.

۱۰ مرکز استان را به تصادف انتخاب کنید و بر اساس میزان بارندگی سالانه در این مراکز، میزان بارندگی سالانه در ایران در این سال را حدوداً به‌دست آورید.

ــ در این جامعه چند متغیر وجود دارد؟ نام ببرید و نوع آن‌ها را مشخص کنید.

ــ آیا میزان بارندگی سالانه در مراکز استان‌ها، به میزان بارندگی سالانه در ایران نزدیک است یا خیر؟

ــ برای اندازه‌گیری تفاوت‌ها از چه معیارهایی استفاده می‌کنید؟

ــ میزان بارندگی به طور متوسط در هر ماه در رشت چقدر است؟

سطر مربوط به بارندگی در استان گیلان (به مرکز رشت) معرف چه متغیری است؟ آن متغیر را معرفی کنید و نوع آن را مشخص کنید.

مقدار بارندگی سالانه مراکز استانها (میلیمتر)

مرکز استان	۱۳۶۵	۱۳۷۰	۱۳۷۲	۱۳۷۳	۱۳۷۴	۱۳۷۵	۱۳۷۶
اراک	۳۴۸	۳۲۰	۳۲۰	۳۵۶	۳۶۷	۳۰۳	۲۵۶
اردبیل	...	...	۳۲۸	۳۴۴	۲۳۰	۳۴۷	۳۴۲
ارومیه	۳۸۶	۲۶۹	۵۳۴	۴۸۶	۳۶۷	۲۶۳	۲۶۹
اصفهان	۱۳۸	۱۱۲	۷۷	۷۹	۲۰۷	۶۷	۲۱۱
اهواز	۲۸۶	۲۹۲	۱۶۸	۲۰۶	۲۳۹	۱۵۲	۵۳۹
ایلام	...	۶۷۶	۷۴۱	۸۶۱	۵۴۹	۵۷۸	۷۱۷
بندرعباس	۲۳۰	۲۴۴	۳۶	۷۰	۵۰۷	۱۶۷	۴۳۷
بوشهر	۱۱۹	۱۷۱	۶۴	۲۸۷	۵۸۸	۲۱۷	۸۰۷
تبریز	۳۲۳	۲۳۶	۳۳۹	۳۳۶	۱۵۸	۲۵۹	۲۰۲
تهران	۲۹۲	۲۶۰	۲۱۹	۲۵۵	۳۱۹	۱۵۶	۲۴۴
خرم‌آباد	۵۹۴	۵۱۶	۵۸۷	۵۸۹	۵۳۷	۴۲۸	۶۰۹
رشت	۱۱۸۶	۱۲۹۸	۱۶۸۵	۱۳۱۹	۱۲۱۰	۱۰۱۵	۱۲۹۷
زاهدان	۴۲	۴۴	۵۴	۵۶	۱۹۰	۷۳	۱۴۹
زنجان	۳۲۱	۲۸۶	۴۲۵	۳۳۵	۲۶۶	۲۴۹	۳۳۱
سمنان	۲۲۲	۱۹۴	۶۷	۱۲۲	۱۹۲	۱۲۸	۱۶۱
سنندج	۵۴۸	۳۸۲	۵۹۹	۵۶۱	۵۱۳	۴۱۳	۵۷۶
شهرکرد	۴۹۲	۴۹۱	۲۸۲	۳۶۸	۳۷۹	۲۵۸	۴۲۳
شیراز	۴۹۵	۲۷۲	۱۵۵	۴۵۱	۵۷۸	۲۰۲	۴۹۸
قزوین	...	...	...	...	...	...	۳۳۶
قم	...	...	...	...	...	۷۸	۱۳۸
کرمان	۱۵۹	۱۱۵	۵۵	۱۷۶	۲۵۶	۱۲۳	۱۷۹
کرمانشاه	۵۶۹	۴۵۱	۶۵۲	۵۵۹	۴۳۳	۳۹۶	۴۸۸
گرگان	...	...	...	...	...	...	۵۳۶
مشهد	۲۵۰	۳۲۶	۳۲۷	۱۶۷	۲۲۴	۱۷۰	۳۴۸
همدان	۴۱۷	۲۹۲	۲۸۸	۴۳۸	۳۱۹	۲۱۳	۳۳۶
یاسوج	۱۲۲۸	۱۲۰۲	۴۱۷	۱۰۶۳	۱۰۲۵	۴۷۴	۱۱۹۸
یزد	۶۴	۶۰	۲۱	۳۴	۱۰۴	۲۹	۴۲

مأخذ : سازمان هواشناسی کشور.

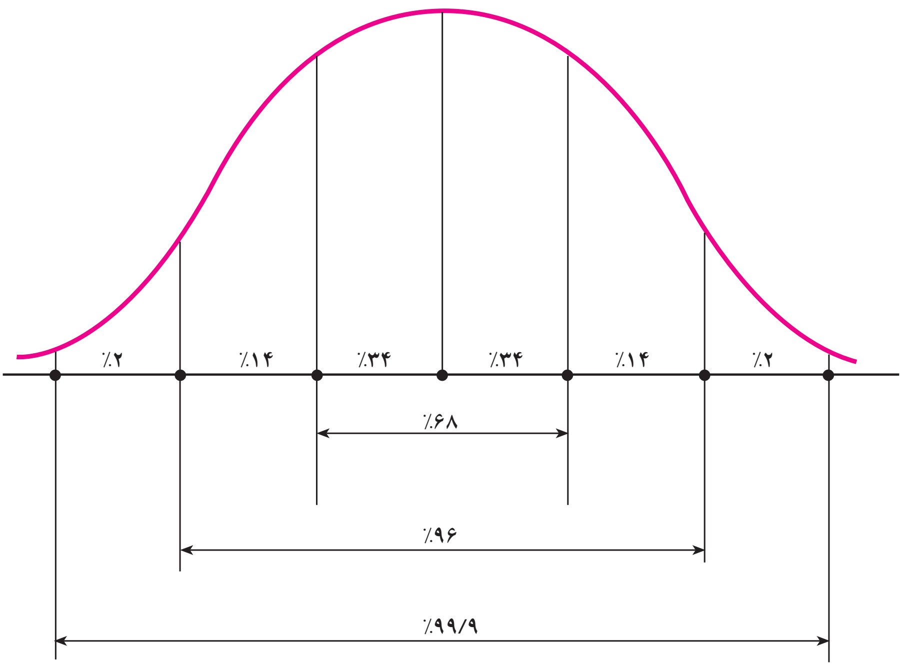
پراکندگی در منحنی نرمال

قبلاً در فصل نمودارها و همچنین در فصل شاخص‌های مرکزی به منحنی نرمال اشاره کردیم.

اگر σ انحراف معیار این جامعه باشد پراکندگی افراد در جامعه به‌صورت زیر است. برای درک بهتر موضوع ابتدا فرض می‌کنیم مثلاً میانگین برابر ۵ و انحراف معیار برابر ۲ باشد در این صورت می‌خواهیم بدانیم چند درصد افراد در فاصله‌ای به مرکز میانگین و به فاصله یک انحراف معیار از آن قرار دارند. درواقع می‌خواهیم بدانیم چند درصد افراد در فاصله (۲+۵ ، ۲−۵) یعنی (۳,۷) قرار دارند.

در این منحنی نشان داده می‌شود که این درصد برابر ۶۸٪ است.

همچنین درصد افرادی که در فاصله‌ای به مرکز میانگین و به طول دو برابر انحراف معیار از میانگین قرار دارند برابر ۹۶٪ است و سرانجام تقریباً ۱۰۰٪ افراد در فاصله‌ای به مرکز میانگین و به‌طول ۳ برابر انحراف معیار قرار دارند. این اطلاعات را به‌طور مختصر در شکل زیر ملاحظه می‌کنید (مقادیر روی نمودار تقریبی هستند.) :



محاسبات شاخص‌های عددی در MINITAB

ــ داده‌های زیر ساعاتی را که ۱۰ دانشجو (انتخاب شده به‌طور تصادفی) صرف تحقیق در طول هفته می‌کنند نشان می‌دهد :

۸, ۰, ۴, ۱۰, ۷, ۲, ۰, ۵, ۱۴, ۷

با استفاده از MINITAB میانگین، میانه، دامنه تغییرات و انحراف معیار این داده‌ها را به‌دست آورید.

شما می‌توانید از Calc و Stat در لیست انتخاب استفاده نمایید.

Stat مقادیر چند شاخص عددی را هم‌زمان ارائه می‌دهد ولی Calc مقدار شاخص‌ها را یکی‌یکی به‌دست می‌آورد. با استفاده از MINITAB WINDOWS برای انتخاب Stat مراحل زیر را دنبال کنید :

قدم اول: داده‌ها را در ستون C1 وارد کنید.

قدم دوم: روی Stat در لیست انتخاب، کلیک کنید.

قدم سوم: روی Basic Statistics کلیک کنید.

قدم چهارم: روی Descriptive Statistics کلیک کنید.

قدم پنجم: C1 را در کار زیر Variable تایپ کنید.

قدم ششم: روی OK کلیک کنید. نتایج و خروجی Minitab روی صفحه نمایان می‌شود.

اگر از فرمان‌های زبانی MINITAB استفاده می‌کنید، با استفاده از SET داده‌ها را در ستون C1 وارد کنید. سپس فرمان زیر را تایپ کنید :

```
MTB > DESCRIBE   C1
```

با اجرای این فرمان خروجی Minitab ظاهر می‌شود.

فعالیت

فعالیت مثال بالا را اجرا کنید و مقادیر به‌دست آمده را در جدول زیر بنویسید.

Descriptive Statistics

Varable	N	Mean	Median	Tr Mean	Stder	SE MEAN
hours	?	?	?	?	?	?

Variable	Min	Max	Q1	Q3
hours	?	?	?	?

همان‌طور که گفته شد می‌توان از دستور Calc نیز استفاده کرد، اگر از Minitab Windows استفاده می‌کنید به‌ترتیب زیر عمل کنید :

قدم اول: داده‌ها را در ستون C1 وارد کنید.

قدم دوم: روی Calc در لیست انتخاب، کلیک کنید.

قدم سوم: روی Column Statistics کلیک کنید.

قدم چهارم: در صفحه Column Statistics روی شاخصی که می‌خواهید محاسبه کنید کلیک کنید. به‌طور مثال برای پیدا کردن میانگین روی Mean کلیک کنید.

قدم پنجم: در کنار Input Variable در کادر، C1 را تایپ کنید.

قدم ششم: اگر می‌خواهید پاسخ را به‌عنوان یک عدد ثابت ذخیره کنید در کادر در کنار Store Result in، K1 را تایپ کنید. اگر از K1 قبلاً استفاده کرده‌اید K2 را تایپ کنید. اگر نمی‌خواهید نتیجه را ذخیره کنید، این مرحله را اجرا نکنید.

قدم هفتم: روی OK کلیک کنید. نتیجه روی صفحه نمایان خواهد شد.

اگر از فرمان‌های زبانی Minitab استفاده می‌کنید، با استفاده از SET داده‌ها را در ستون C1 وارد کنید. سپس فرمان زیر را تایپ کنید :

MTB > Mean C1

در پاسخ، Minitab خروجی زیر را ارائه می‌کند :

Column Mean
Mean of C1 = ...

به‌طور متشابه می‌توانید میانه، دامنه تغییرات و انحراف معیار را با کلیک کردن روی شاخص مورد نظر محاسبه کنید و فرمان‌های زبانی Minitab برای شاخص‌های فوق عبارتند از :MEDIAN C1، RANGE C1، STANDARD DEVIATION که باید پس از PROMPT تایپ شوند.

نمودار جعبه‌ای

داده‌های زیر، درآمد هفتگی نمونهٔ ۱۲ تایی از خانواده‌ها را نشان می‌دهد.

۵۲۰۰۰	۲۲۰۰۰	۶۰۰۰۰	۳۲۰۰۰	۱۷۰۰۰	۲۳۰۰۰
۴۶۰۰۰	۲۷۰۰۰	۹۲۰۰۰	۴۲۰۰۰	۳۸۰۰۰	۲۹۰۰۰

با استفاده از Minitab نمودار جعبه‌ای این داده‌ها را رسم کنید. اگر از Minitab Windows

استفاده می‌کنید، مراحل زیر را دنبال کنید :

قدم اول: داده‌ها را در ستون C1 وارد کنید.

قدم دوم: روی GRAPH در لیست انتخاب کلیک کنید.

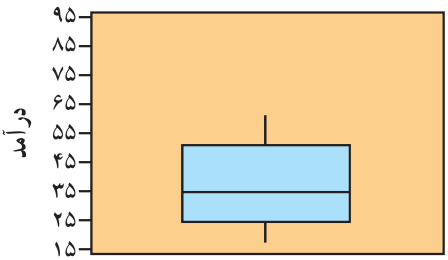
قدم سوم: روی BOXPLOT کلیک کنید.

قدم چهارم: C1 را در کادر زیر Y تایپ کنید.

قدم پنجم: روی OK کلیک کنید. نمودار جعبه‌ای مورد نظر نمایان می‌شود. اگر از فرمان‌های زبانی Minitab استفاده می‌کنید ابتدا با استفاده از فرمان SET داده‌ها را در ستون C1 وارد کنید. سپس فرمان زیر را تایپ کنید :

```
MTB > Boxplot   C1
```

با اجرای دستورات فوق نمودار جعبه‌ای زیر رسم می‌شود :



تمرین

۱ــ نمودار ساقه و برگ زیر موجود است.

ــ میانگین، انحراف معیار این داده‌ها را با استفاده از Minitab به‌دست آورید.

ــ نمودار جعبه‌ای این داده‌ها را با استفاده از Minitab رسم کنید.

ساقه	برگ
۰	۳۶۹
۱	۲۸۵۱۰۵
۲	۵۱۶
۴	۱
۵	
۶	۲

و بدان! کسی که گنجینه‌های آسمان و زمین در دست اوست تو را در دعا رخصت داده و پذیرفتن دعایت را برعهده نهاده و تو را فرموده از او خواهی تا به تو دهد، و از او طلبی تا تو را بیامرزد. و میان تو و خود کسی را نگمارده تا تو را از وی بازدارد و از کسی ناگزیرت نکرده که نزد او برایت میانجی‌گری آرد، و اگر گناه کردی از توبه‌ات منع ننموده و در کیفرت شتاب نفرموده و چون بدو باز گردی سرزنشت نکند، و آنجا که رسوا شدنت سزاست پرده‌ات را ندرد؛ و در پذیرفتن توبه بر تو سخت نگرفته و حساب گناهت را نکشیده، و از بخشایش نومیدت نگردانیده، بلکه بازگشتت را از گناه نیک شمرده و هر گناهت را یکی گرفته و هر کار نیکویت را ده به حساب آورده و در بازگشت را برایت بازگذارده و چون بخوانیش آوایت را شنود، و چون راز خود را با او درمیان نهی آن را داند، پس حاجت خود بدو نمایی و آنچه در دل داری پیش او بگشایی، و از اندوه خویش بدو شکایت کنی و خواهی تا غم تو را گشاید و در کارها یاری‌ات نماید و از گنجینه‌های رحمت او آن را خواهی که بخشیدنش از جز او نیاید: از افزودن مدت زندگانی و تندرستی‌ها و در روزی‌ها فراوانی.

فصل هشتم

(تدریس این فصل اختیاری است)

همبستگی و رگرسیون

شهاب برای پروژه کلاس خود مایل بود تأثیر سطح تحصیلات والدین را بر موفقیّت تحصیلی فرزندان مطالعه کند. در انجام این پروژه او با یک مشکل جدی روبرو بود و آن اینکه جامعه‌ای که با آن سروکار داشت شامل یک متغیر نبود و او نمی‌دانست که در این شرایط چگونه باید عمل کند. از این رو به معلم مراجعه کرد و معلم به او پاسخ داد : جامعه‌ای که می‌خواهی مطالعه کنی یک جامعه با دو متغیّر است که موضوع درس امروز ماست.

ما در بررسی این قبیل جوامع نه تنها کارهایی را که تا بحال آموخته‌ایم به کار می‌بریم، بلکه به دو سؤال نیز سعی می‌کنیم پاسخ دهیم و آن اینکه این متغیّرها چقدر به‌هم وابسته‌اند و دیگر آنکه اگر مقدار یکی از متغیّرها را بدانیم آیا می‌توانیم مقدار متغیّر دیگر را «**حدس**» بزنیم.

جامعه‌های چند متغیّری

تاکنون جوامعی را بررسی کرده‌ایم که در آن‌ها فقط یک متغیّر وجود داشت. شما درباره قد همکلاسی‌های خود مطالعه کرده‌اید. نمودارهای مربوطه، میانگین و واریانس آن را تعیین کرده‌اید. همین کار را درباره وزن همکلاسی‌های خود جداگانه انجام داده‌اید. ولی هیچ‌گاه این دو متغیّر را با هم مطالعه نکرده‌اید. از تأثیرات آن‌ها بر روی یکدیگر اطلاعی ندارید، هنوز نمی‌دانید که اگر قد همکلاسی خود را بدانید آیا می‌توانید وزن او را «حدس» بزنید، یعنی به‌طور تقریبی بگویید که وزن او چقدر است؟ شما نمی‌دانید که آیا افراد قد بلندتر وزن بیشتری دارند یا خیر، تغییرات این دو متغیّر با هم چگونه است؟ تنها کاری که کرده‌اید در یکی از تمرین‌ها تعیین کرده‌اید که از این دو متغیّر کدامیک تغییرات بیش‌تری در مقایسه با دیگری دارد.

می‌توانیم مسائلی را مطرح کنیم که در آن‌ها تعداد متغیّرها از دو تا هم بیشتر باشد. مثلاً در کشاورزی میزان محصول، میزان آب، میزان کود، نوع کود، میزان آفتاب و نوع خاک مؤثرند. ملاحظه می‌شود که در مسائل کشاورزی حداقل با ۶ متغیّر سروکار داریم. بررسی این قبیل مسائل و

دستیابی به نوع تأثیر این متغیّرها بر روی یکدیگر جزء مسائل عمده آماری است و معمولاً در بخش تحقیقات کشاورزی با این قبیل متغیّرها سروکار دارند. ما در این فصل هم به جهت سادگی و هم به‌جهت آنکه بسیاری از مسائل را می‌توان به جامعه‌هایی با دو متغیّر تجزیه کرد به بررسی جامعه‌های دو متغیّری می‌پردازیم.

نمودار پراکنش

داده‌هایی را که تاکنون جمع‌آوری می‌کردید از نظر بگذرانید. به هر فرد یک عدد با یک صفت نسبت داده می‌شد، ولی اکنون به هر فرد دو مقدار نسبت داده می‌شود، زیرا ما دو متغیّر داریم. برای هر متغیّر یک مقدار اندازه گرفته می‌شود. برای آنکه مقادیر مربوط به متغیّرها اشتباه نشود، داده‌ها را به صورت زوج مرتب می‌نویسیم. مثلاً اگر موضوع مورد مطالعه قد و وزن است و شما قد و وزن یکی از همکلاسی‌های خود را اندازه گرفته‌اید، داده‌ای به صورت مرتب شدهٔ (۵۲ و ۱۶۷) خواهید داشت که قد این همکلاسی را برحسب سانتیمتر برابر ۱۶۷ و وزن او را بر حسب کیلو برابر ۵۲ است.

نمایش (۵۲ و ۱۶۷) را یک «زوج مرتب» می‌نامند که نشان می‌دهد عددی که اوّل از سمت چپ نوشته شده نمایانگر قد و عددی که دوّم نوشته شده نمایانگر وزن دانش‌آموز است.

فعالیت

قد و وزن ۱۰ نفر از همکلاسی‌های خود را اندازه بگیرید و داده‌های حاصل را به صورت زوج‌های مرتب نمایش دهید.



البته می‌توان داده‌ها را به صورت مرتب در دو ستون و یا دو سطر کنار هم نیز نشان داد. در یک مطالعه آماری خواسته‌اند تأثیر نوعی رژیم غذایی را روی موش‌های آزمایشگاهی بررسی کنند، از این رو وزن اولیه هر موش را اندازه گرفتند و پس از اعمال رژیم غذایی در مدت معینی، مجدداً وزن آن‌ها را اندازه گرفتند و داده‌ها عبارتند از :

(۲،۴) و (۳،۵) و (۱،۷) و (۲،۳) و (۱،۸)

که مؤلفه اول بیانگر وزن اولیه و مؤلفه دوم بیانگر وزن همان موش پس از اعمال رژیم غذایی است.

نوع دیگری که می‌توان این داده‌ها را نشان داد چنین است :

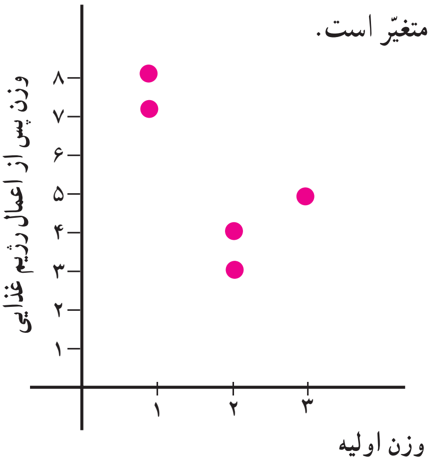
شمارهٔ موش‌ها	۱	۲	۳	۴	۵
وزن اولیه	۱	۲	۱	۳	۲
وزن پس از اعمال رژیم غذایی	۸	۳	۷	۵	۴

ممکن بود این اطلاعات را به صورت ستونی هم بنویسیم.

شمارهٔ موش‌ها	وزن اولیه	وزن پس از رژیم
۱	۱	۸
۲	۲	۳
۳	۱	۷
۴	۳	۵
۵	۲	۴

در هر صورت ثبت داده‌ها باید به گونه‌ای باشد که اولاً داده‌های مربوط به یک فرد در کنار هم باشند و ثانیاً مشخص باشد که کدام مقدار مربوط به کدام متغیّر است.

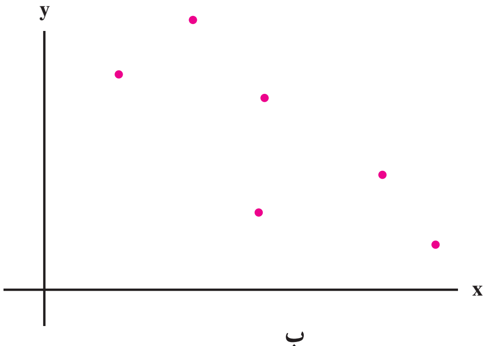
حال که داده‌ها را قرار شد به صورت زوج مرتب نشان دهیم، می‌توانیم آن‌ها را به صورت نقاطی در صفحه نمایش دهیم. شکل روبرو داده‌های مربوط به وزن موش‌ها را نشان می‌دهد. این نمودار را *نمودار پراکنش* می‌گویند.



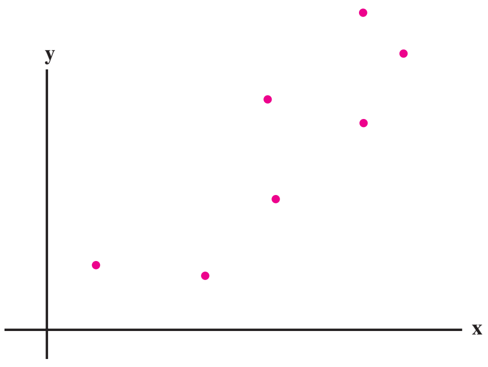


نمودار پراکنش **مجموعه‌ای از نقاط در صفحه محورهای مختصات است که از رسم داده‌های مربوط به اندازه‌گیری در جامعه‌های دو متغیّری به دست آمده است.**

مطالعه این نمودار می‌تواند اطلاعاتی درباره جامعه در اختیار ما بگذارد.
مثلاً نمودارهای زیر را در نظر بگیرید.

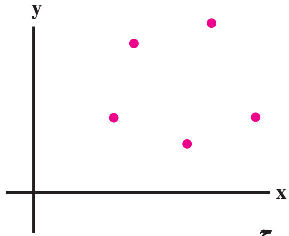


ب

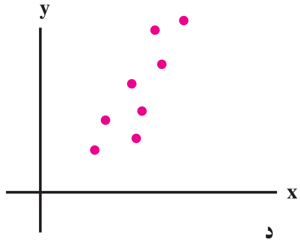


الف

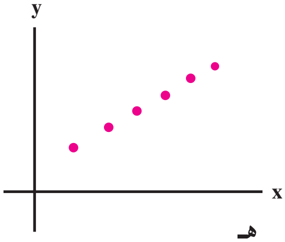
در نمودار «الف» وقتی مقادیر روی محورxها افزایش می‌یابد، مقادیر روی محورyها «عموماً» افزایش می‌یابد امّا در نمودار« ب »برعکس است با افزایش مقادیر روی محورxها، مقادیر روی محورyها «عموماً» کاهش می‌یابد.



ج



د



هـ

در نمودار «ج» تغییرات در مقدار x ها هیچ گونه اطلاعی درباره تغییرات y نمی‌دهد، یعنی رفتار مقادیر x رفتار مقادیر y را مشخص نمی‌کند. نمودار «د» علاوه بر آنکه نشان می‌دهد دو متغیّر با هم افزایش و یا کاهش می‌یابند، با داشتن مقادیر x با دقت بیشتری می‌توان y را حدس زد. دقت حدس زدن مقادیر y در این نمودار از دقت حدس در نمودار «الف» و «ب» بیشتر است. در نمودار «هـ» حدس مقادیر y با دقت بیشتری انجام می‌شود. برای روشن‌تر شدن این مفاهیم فعالیت زیر را انجام دهید.

فعالیت

نمودار پراکنش مربوط به داده‌های زیر را رسم کنید.

(۵،۶) و (۴،۵) و (۳،۴) و (۲،۳) و (۱،۲)

اگر نمودار را دقیق رسم کرده باشید متوجه خواهید شد که این نقاط روی یک خط راست قرار دارند، می‌توانید معادله آن خط را بیابید.

حال مطابق دستور العمل زیر مؤلفه دوم هر یک از این داده‌ها را تغییر دهید : برای هر نقطه سکه‌ای را پرتاب می‌کنیم اگر شیر آمد به y یک واحد اضافه می‌کنیم و اگر خط بیاید از y یک واحد کم می‌کنیم. با تغییرات حاصل شده نمودار پراکنش را رسم کنید.
این فعالیت را تکرار کنید منتها این بار مقادیری را که اضافه و یا کم می‌کنید تغییر دهید.

چه تغییری در نمودار پراکنش مشاهده می‌کنید؟

وابستگی x و y چگونه تغییر می‌کند؟

در این فعالیت می‌بینید که هر چقدر وابستگی خطی y و x بیشتر باشد مقدار y حاصل از داده‌ها به مقدار y حاصل از معادله خط نزدیکتر است. یعنی با داشتن x مقدار y را با خطای کمتری می‌توانید تعیین کنید.

شما می‌توانید فعالیت بالا را به‌جای خط با هر منحنی دیگری تکرار کنید. ما در این فصل فقط با خط کار می‌کنیم.

تمرین

۱ ــ تعداد ساعات صرف شده و نمره کسب شده در آزمون فارسی برای هر دانش‌آموز کلاس دوم یک دبیرستان به شکل زوج مرتب در زیر آورده شده است.

(۱۴ , ۴) و (۱۰ , ۴) و (۱۴ , ۳) و (۱۰ , ۳) و (۱۰ , ۲)

(۱۸ , ۶) و (۱۶ , ۶) و (۱۲ , ۶) و (۱۶ , ۵) و (۱۴ , ۵)

(۱۸ , ۸) و (۱۶ , ۸) و (۲۰ , ۷) و (۱۸ , ۷) و (۱۴ , ۷)

نمودار پراکنش را برای نشان دادن ارتباط بین تعداد ساعات مطالعه و نمرهٔ آزمون رسم کنید. چه نتیجه‌ای می‌گیرید.

۲ ــ نمودار زیر قیمت و عمر تعدادی اتومبیل را نشان می‌دهد.

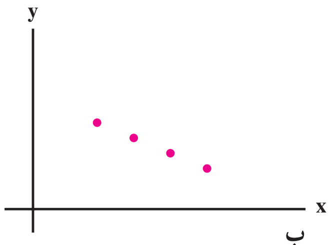


آیا ارتباطی بین این دو متغیر وجود دارد؟ توضیح دهید.

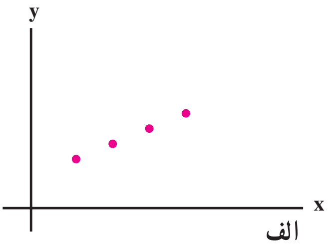


همبستگی

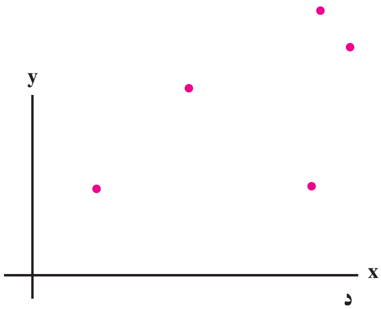
در فعالیت‌های انجام شده دیده‌اید که پس از تغییراتی که در داده‌ها داده‌اید در برخی از آن‌ها نقاط به شکل یک خط نزدیکتر و در برخی دیگر دورتر می‌شوند. اگر نقاط به شکل یک خط نزدیکتر باشند به کمک x با خطای کمتری می‌توان y را تعیین کرد و برعکس هرچقدر نقاط از خط دورتر باشند خطای تعیین y بیشتر خواهد بود. هر چقدر خطای تعیین y به کمک x کمتر باشد می‌نویسیم دو متغیر همبستگی خطی بیشتری دارند مثلاً در شکل‌های زیر



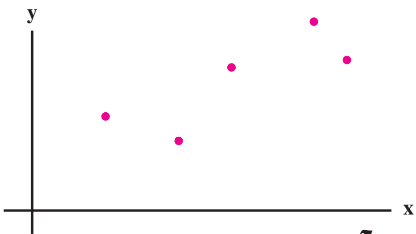
ب



الف



د



ج

در «الف» و«ب» بدون خطا می‌توانیم y را به کمک خطی که از این نقاط می‌گذرد تعیین کنیم. پس همبستگی بین x و y خیلی زیاد است. ولی در شکل «ج» نمی‌توانیم خطی را تعیین کنیم که تمام نقاط روی آن باشند. به نظر می‌رسد این نقاط به نحوی از خط راست دور شده‌اند پس همبستگی آن به اندازه همبستگی در شکل‌های «الف» و «ب» نیست. ولی باز خطای تعیین y به کمک x زیاد نیست. امّا این خطا در شکل «د» خیلی زیاد است.

ما معیاری می‌خواهیم که این همبستگی را برای ما اندازه‌گیری کند. دستوری که برای محاسبه همبستگی معرفی شده است با نام ضریب همبستگی و با نماد r، به صورت زیر است :

$$r = \frac{\dfrac{(\text{مجموع حاصلضرب } x\text{ها و } y\text{ها})}{n} - (\text{میانگین } x\text{ها})(\text{میانگین } y\text{ها})}{(\text{انحراف معیار } x\text{ها})(\text{انحراف معیار } y\text{ها})}$$

در زیر داده‌هایی مربوط به وزن اولیه و بعد از رژیم موش‌ها آمده است.

شماره موش‌ها	۱	۲	۳	۴	۵	جمع
وزن اولیه (x)	۱	۱	۲	۳	۲	۹
وزن بعد از اعمال رژیم غذایی (y)	۸	۳	۷	۶	۴	۲۸
حاصلضرب xها و yها	۸	۳	۱۴	۱۸	۸	۵۱

ضریب همبستگی بین وزن اولیه و وزن بعد از اعمال رژیم غذایی را محاسبه کنید.

در این مسأله $n = ۵$ و xها را وزن اولیه وyها را وزن بعد از اعمال رژیم غذایی گرفته‌ایم.

ابتدا به محاسبهٔ میانگین x و y می‌پردازیم :

$$\bar{x} = \frac{۹}{۵} = ۱/۸ \quad , \quad \bar{y} = \frac{۲۸}{۵} = ۵/۶ \ ,$$

$$\text{واریانس xها} = \frac{(۱-۱/۸)^۲ + (۱-۱/۸)^۲ + (۲-۱/۸)^۲ + (۳-۱/۸)^۲ + (۲-۱/۸)^۲}{۵}$$

$$= \frac{۰/۶۴ + ۰/۶۴ + ۰/۰۴ + ۱/۴۴ + ۰/۰۴}{۵} = \frac{۲/۸}{۵} = ۰/۵۶$$

$$\text{واریانس yها} = \frac{(۸-۵/۶)^۲ + (۳-۵/۶)^۲ + (۷-۵/۶)^۲ + (۶-۵/۶)^۲ + (۴-۵/۶)^۲}{۵}$$

$$= \frac{۵/۷۶ + ۶/۷۶ + ۱/۹۶ + ۰/۱۶ + ۲/۵۶}{۵} = \frac{۱۷/۲}{۵} = ۳/۴۴$$

$$\frac{\text{مجموع حاصلضرب xها و yها}}{n} = \frac{۵۱}{۵} = ۱۰/۲$$

بنابراین :

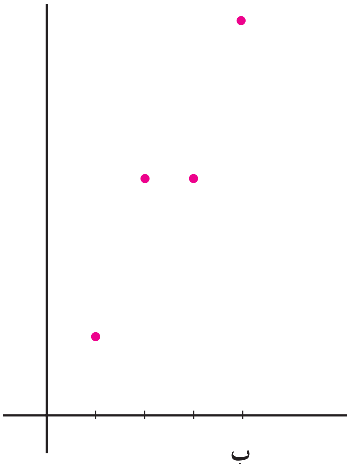
$$\text{ضریب همبستگی} = \frac{۱۰/۲ - (۱/۸)(۵/۶)}{(\sqrt{۰/۵۶})(\sqrt{۳/۴۴})} \approx \frac{۰/۱۲}{۱/۳۹}$$

$$\approx ۰/۰۹$$

حال این سؤال پیش می‌آید که آیا ضریب همبستگی ۰/۰۹ نشان از همبستگی زیاد بین x و y است و یا خیر. برای آنکه بتوانیم به این سؤال پاسخ دهیم ضریب همبستگی را در برخی حالت‌های زیر حساب می‌کنیم :

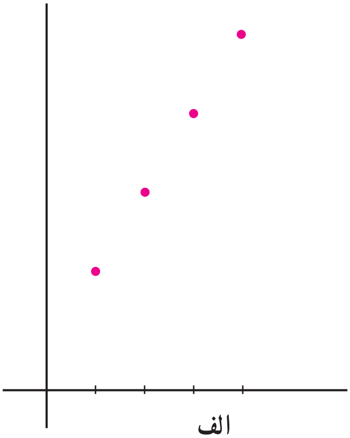


x	۱	۲	۳	۴
y	۲	۶	۶	۱۰



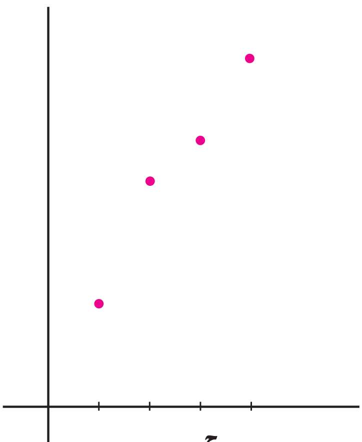
ب

x	۱	۲	۳	۴
y	۳	۵	۷	۹



الف

x	۱	۲	۳	۴
y	۲/۵	۵/۵	۶/۵	۸/۵



ج

در شکل «الف» نقاط کاملاً روی خط قرار دارند پس باید بالاترین ضریب همبستگی را داشته باشند.

در شکل «ب» نقاط از خط دورتر شده‌اند پس باید ضریب همبستگی آن از ضریب همبستگی «الف» کمتر باشد.

در شکل «ج» نقاط از خط دور شده‌اند ولی نه به اندازه دوری حالت «ب»، پس باید ضریب همبستگی در این حالت از ضریب همبستگی در حالت «ب» بیشتر ولی از ضریب همبستگی حالت «الف» کمتر باشد.

حال با محاسبه صحت مطالب بالا را آزمایش می‌کنیم.

ضریب همبستگی در حالت «الف»

$\bar{x}=\frac{۱۰}{۴}=۲/۵ \quad , \quad \bar{y}=\frac{۲۴}{۴}=۶$

واریانس x $=\frac{۲/۲۵+۰/۲۵+۰/۲۵+۲/۲۵}{۴}=\frac{۵}{۴}=۱/۲۵$

واریانس y $=\frac{۹+۱+۱+۹}{۴}=\frac{۲۰}{۴}=۵$

مجموع حاصلضرب xها و yها $=۳+۱۰+۲۱+۳۶=۷۰$

ضریب همبستگی $=\frac{\frac{۷۰}{۴}-(۲/۵)(۶)}{\sqrt{۱/۲۵}\times\sqrt{۵}}=\frac{۱۷/۵-۱۵}{\sqrt{۶/۲۵}}=\frac{۲/۵}{۲/۵}=۱$

ضریب همبستگی در حالت «ب»

$\bar{x}=۲/۵ \quad , \quad \bar{y}=\frac{۲۴}{۴}=۶$

واریانس x $=۱/۲۵$,

واریانس y $=\frac{۱۶+۰+۰+۱۶}{۴}=۸$

مجموع حاصلضرب xها و yها $=۲+۱۲+۱۸+۴۰=۷۲$

ضریب همبستگی $=\frac{\frac{۷۲}{۴}-(۲/۵)(۶)}{\sqrt{۱/۲۵}\times\sqrt{۸}}=\frac{۳}{\sqrt{۱۰}}\approx۰/۹۴۸۶$

ضریب همبستگی در حالت «ج»

$\bar{x}=۲/۵$ ، واریانس x $=۱/۲۵$

میانگین y $=\frac{۲۳}{۴}=۵/۷۵$

واریانس y $=\frac{(۳/۲۵)^{۲}+(۰/۲۵)^{۲}+(۰/۷۵)^{۲}+(۲/۷۵)^{۲}}{۴}$

$=\frac{۱۸/۷۵}{۴}=۴/۶۸۷۵$

$\frac{\text{مجموع حاصلضرب xها و yها}}{۴}=\frac{۲/۵+۱۱+۱۹/۵+۳۴}{۴}=\frac{۶۷}{۴}=۱۶/۷۵$

ضریب همبستگی $=\frac{۱۶/۷۵-(۲/۵)(۵/۷۵)}{\sqrt{۱/۲۵}\ \sqrt{۴/۶۸۷۵}}=\frac{۲/۳۷۵}{\sqrt{۵/۸۵۸}}\approx۰/۹۸۱۲$

با مقایسهٔ محاسبات در حالت‌های مختلف ملاحظه می‌شود که نتیجهٔ محاسبات با انتظاری که از ضریب همبستگی داریم هماهنگی دارد.

البته ما هنوز به سؤالی که بعد از مثال از خود پرسیده بودیم پاسخ نداده‌ایم. سؤال این بود که ضریب همبستگی به اندازه ۰/۰۹ چه معنایی دارد. آیا این نشان از همبستگی زیاد است یا همبستگی کم. ما در این قسمت وارد تحلیل اندازه ضریب همبستگی نمی‌شویم فقط اضافه می‌کنیم که در عرف محاسبات آماری ۰/۰۹ ضریب همبستگی بالایی نیست.

ما تا بحال ضریب همبستگی‌هایی که محاسبه کردیم مثبت بودند، ممکن است سؤال شود آیا ممکن است ضریب همبستگی منفی هم باشد؟ پاسخ به این سؤال مثبت است.

داده‌های زیر را در نظر بگیرید.

x	۱	۲	۳	۴	۵
y	۴	۳	۲	۱	۰

ضریب همبستگی این داده‌ها را محاسبه می‌کنیم.

البته با رسم نمودار پراکنش متوجه می‌شویم که این داده‌ها روی خط $y=-x+۵$ قرار دارند.

$$\bar{x}=\frac{۱+۲+۳+۴+۵}{۵}=۳$$

$$\bar{y}=\frac{۴+۳+۲+۱}{۵}=۲$$

$$\text{واریانس } x=\frac{۴+۱+۰+۱+۴}{۵}=۲$$

$$\text{واریانس } y=\frac{۴+۱+۰+۱+۴}{۵}=۲$$

$$\frac{\text{مجموع حاصلضرب } x\text{ها و } y\text{ها}}{۵}=\frac{۴+۶+۶+۴+۰}{۵}=\frac{۲۰}{۵}=۴$$

$$\text{ضریب همبستگی}=\frac{۴-(۳)(۲)}{\sqrt{۲}\times\sqrt{۲}}=\frac{-۲}{۲}=-۱$$



حتماً متوجه شده‌اید در شرایطی که تعیین y با کمترین خطا انجام شد (مقادیر y و x روی یک خط قرار دارند) ضریب همبستگی ۱ و یا ۱− به دست آمد، در سایر شرایط ضریب همبستگی بین این دو مقدار است. در این مرحله ما قادر به اثبات ویژگی‌های ضریب همبستگی نیستیم فقط به ذکر آن‌ها می‌پردازیم.

ــ ضریب همبستگی همواره بین ۱− و ۱+ قرار دارد.

ــ اگر مقادیر x و y روی خطی با شیب زاویه مثبت باشد، ضریب همبستگی x و y برابر ۱+ است.

ــ اگر مقادیر x و y روی خطی با ضریب منفی زاویه قرار گیرد، ضریب همبستگی x و y برابر ۱− است.

ــ اگر ضریب همبستگی مثبت باشد با افزایش x ، مقادیر y نیز عموماً افزایش می‌یابد و بالعکس.

ــ اگر ضریب همبستگی منفی باشد با افزایش x ، مقادیر y عموماً کاهش می‌یابد و بالعکس.

ــ هر چقدر ضریب همبستگی به ۰ نزدیکتر باشد تعیین y با خطای بیشتری انجام می‌شود.

ــ هر چقدر ضریب همبستگی به ۱ یا ۱− نزدیکتر باشد تعیین y با خطای کمتری انجام می‌شود.

تمرین

۱ ــ ضریب همبستگی قد و وزن همکلاسی‌های خود را حساب کنید.

۲ ــ ضریب همبستگی بین نمرهٔ درس آمار و مدلسازی و نمره یکی دیگر از درس‌ها را در کلاس حساب کنید.

۳ ــ ضریب همبستگی بین معدل دانش‌آموزان و تعداد افراد خانواده آن‌ها را حساب کنید.

۴ ــ گمان می‌کنید ضریب همبستگی بین کدام دو درس در کلاس بیش از همه است؟ درستی گمان خود را چگونه تحقیق می‌کنید؟

۵ ــ در تمرین‌های ۱ تا ۴ به سؤال‌های زیر پاسخ دهید :

ــ آیا می‌توانید علامت ضریب همبستگی را توجیه کنید؟

ــ در کدام یک از موارد ضریب همبستگی به ۰ نزدیکتر است؟

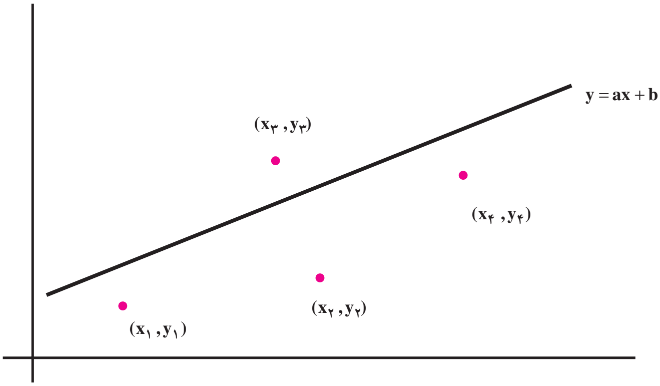
ــ آیا نتایج حاصل از تحلیل ضریب همبستگی‌ها معقول به نظر می‌رسد؟



خط رگرسیون

تا بحال صحبت از این بود که هر چقدر نقاط به خط نزدیک‌تر باشند، ضریب همبستگی از صفر دورتر است و به ۱ یا ۱− نزدیک‌تر خواهد بود. ما اگر معادلهٔ خطی را که از آن صحبت می‌کنیم داشته باشیم و بدانیم ضریب همبستگی هم در حد قابل قبولی است آن گاه می‌توانیم با داشتن x مقدار y را از روی خط تعیین کنیم و تا «حدودی» مطمئن باشیم که خطای زیادی مرتکب نشده‌ایم. امّا در مسائل عملی ما این خط را نداریم، در این مبحث می‌خواهیم روش یافتن این خط را بیان کنیم.

روش کمترین مربعات: فرض کنید نمودار پراکنش به صورت زیر باشد.



می‌خواهیم خطی به معادله $y = ax + b$ از بین این نقاط بگذرانیم به قسمتی که به همه نقاط «نزدیک» باشد. ما ذیلاً توضیح می‌دهیم که نزدیکی این خط به همه نقاط به چه معنی است؟

ما از این خط می‌خواهیم برای به دست آوردن yها استفاده کنیم، مثلاً به جای $y_۱$ می‌خواهیم از مقدار $ax_۱ + b$ که از روی خط حاصل شده است استفاده کنیم، در این صورت مقدار خطایی که در این جابجایی مرتکب شده‌ایم عبارت است از $(ax + b - y_۱)$ و همین طور برای سایر نقاط می‌توانیم

این خطاها را حساب کنیم. از آنجایی که تعداد این خطاها ممکن است مثبت و یا منفی باشد، با به توان دو رساندن آن‌ها همگی را مثبت خواهیم کرد، در واقع می‌خواهیم خط را چنان تعیین کنیم تا مجموع مجذور خطاها می‌نیمم شود، یعنی می‌خواهیم a و b را چنان به دست آوریم که عبارت :

$$(ax_۱+b-y_۱)^۲+(ax_۲+b-y_۲)^۲+(ax_۳+b-y_۳)^۲+\cdots$$

کمترین مقدار ممکن را بگیرد. این روش را روش ***کمترین مربعات*** می‌گویند و ***خط*** حاصل را ***خط رگرسیون y نسبت به x*** می‌گویند. ما بدون آنکه وارد روش‌های ریاضی محاسبه a و b شویم فقط به ذکر دستور محاسبه آن‌ها می‌پردازیم.

محاسبه a و b به روش کمترین مربعات

$$a=r\frac{\sigma_y}{\sigma_x}=\text{ضریب همبستگی}\times\frac{\text{انحراف معیار } y}{\text{انحراف معیار } x}$$

$$b=\bar{y}-a\bar{x}$$

حال خط رگرسیون y نسبت به x را در مسأله وزن موش‌ها به‌دست می‌آوریم. قبلاً دیدیم که :

$$\bar{x}=۱/۶\quad,\quad \bar{y}=۵/۴$$

$$\text{انحراف معیار } x=\sqrt{۰/۶}\qquad،\qquad \text{انحراف معیار } y=\sqrt{۳/۴۴}$$

$$r=۰/۱۱$$

بنابراین :

$$a=۰/۱۱\frac{\sqrt{۳/۴۴}}{\sqrt{۰/۶}}\approx ۰/۲۶۳$$

$$b=۵/۴-۱/۶=۴/۹۷$$

پس معادلهٔ خط رگرسیون عبارت است از :

$$y=۰/۲۶۳x+۴/۹۷$$

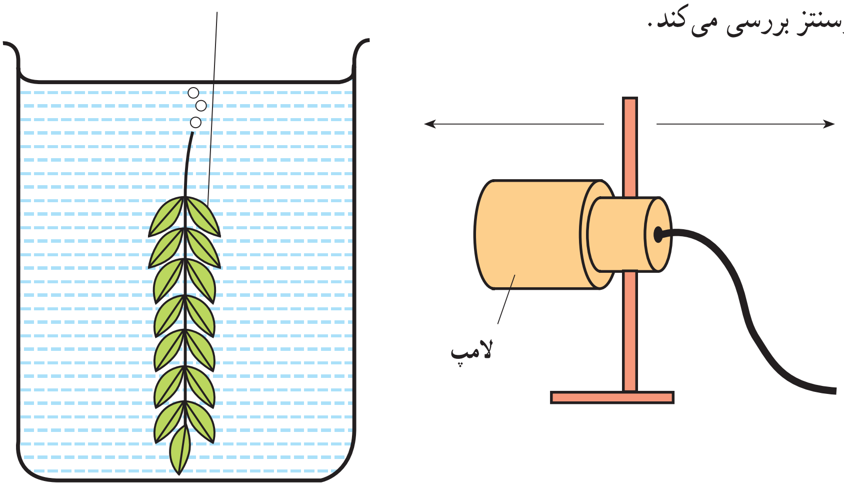
حال اگر وزن اولیه موشی ۱/۵ باشد، پس از اعمال رژیم غذایی وزن آن حدوداً چقدر خواهد بود؟ برای پاسخ به این سؤال از خط رگرسیون y نسبت به x استفاده می‌کنیم داریم $x=۱/۵$ و پس

$$y=(۰/۲۶۳)(۱/۵)+۴/۹۷=۵/۳۶$$

خط رگرسیون ابزاری است برای پیش‌بینی مقدار یک متغیر برحسب متغیری که به آن وابسته است.

تمرین

۱ــ مهرناز آزمایشی را تدارک دیده است که اثر نور را بر سرعت تهیه غذا توسط گیاه ازطریق فتوسنتز بررسی می‌کند.



مهرناز لامپ را در ۱۰۰ متری گیاه قرار داده و تعداد حباب‌هایی را که توسط گیاه در یک دقیقه تولید شد، شمرد. سپس لامپ را نزدیک‌تر کرده و تعداد حباب‌های ایجاد شده توسط گیاه در دقیقه را شمرد وجدول زیر را تهیه نمود.

تعداد حباب‌ها در دقیقه	فاصله لامپ از گیاه برحسب cm
۱۰	۱۰۰
۲۰	۸۰
۳۲	۶۰
۳۷	۴۰
۳۷	۲۰

ــ نمودار پراکنش این داده‌ها را رسم کنید.

ــ بدون محاسبه و به‌طور تقریبی، خط رگرسیون را رسم کنید.

ــ اگر لامپ را به ۱۰ سانتی‌متری گیاه منتقل کنیم، تعداد حباب‌ها در دقیقه را حدس بزنید.

ــ با استفاده از نتیجه حاصل شده جمله زیر را کامل کنید :

همان‌طور که فاصله لامپ ....، سرعت فتوسنتز ....

۲ــ نمودار پراکنش نتایج امتحانات درس ریاضی و درس فیزیک ۲۰ دانش‌آموز که به شرح زیر آمده است را رسم کنید.

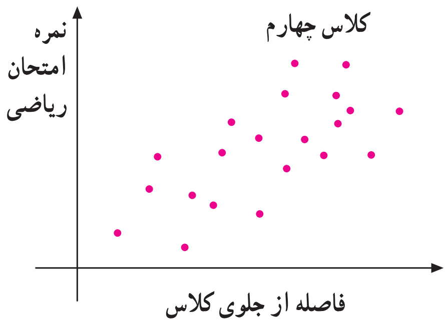
دانش‌آموز	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱
نمره ریاضیات	۱۸	۸	۱۲	۱۸	۱۶	۶	۱۰	۶	۱۶	۱۰	۱۲
نمره فیزیک	۱۶	۱۰	۸	۱۸	۱۴	۱۰	۱۰	۶	۱۰	۱۴	۱۰

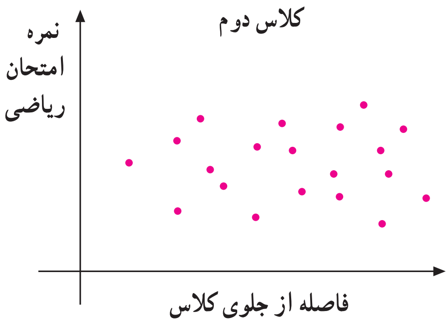
دانش‌آموز	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰
نمره ریاضیات	۱۲	۱۰	۱۶	۱۸	۱۲	۶	۱۴	۱۴	۱۲
نمره فیزیک	۱۶	۱۲	۱۶	۱۲	۱۲	۸	۱۲	۱۴	۱۶

ــ معادله خط رگرسیون را به‌دست آورید.

ــ آیا ارتباطی بین نتایج دو تست وجود دارد؟

۳ــ برخی از دانش‌آموزان تصور می‌کنند هرکس که در ردیف‌های جلوی کلاس می‌نشینند، نمرات بهتری دردرس ریاضی کسب می‌کند. یکی از معلمین تصمیم گرفت این موضوع را در دو کلاس خود مورد بررسی قرار دهد. نمودارهای زیر نتیجهٔ این بررسی است.





درستی نظر دانش‌آموزان را در دو کلاس مورد بررسی قرار دهید. دو نمودار فوق را تفسیر کنید.

۴ــ آشپزی زمان لازم برای پختن گوشت غذا به وزن‌های مختلف را در جدول صفحه بعد ثبت نموده است.

ــ روی محور افقی با درنظر گرفتن یک سانتی‌متر نشان‌گر یک کیلوگرم و روی محور عمودی با درنظر گرفتن هر سانتی‌متر نشان‌گر ۳۰ دقیقه، نمودار پراکنش را برای نشان‌دادن ارتباط بین وزن گوشت و زمان لازم برای پختن آن رسم کنید.

ــ خط رگرسیون را رسم کنید.

ــ با استفاده از خط رگرسیون زمان لازم را برای پختن ۳ کیلوگوشت تخمین بزنید.

وزن گوشت (kg)	زمان برحسب دقیقه
۲	۱۳۰
۳/۵	۲۰۰
۵	۲۱۵
۵/۵	۲۸۰
۶/۵	۲۷۰

۵ ــ خانم خانه‌داری از فرزندش که درس آمار خوانده بود خواست تا روی اندازهٔ سیب‌زمینی‌های داخل کیسه‌های ۲ کیلویی بررسی انجام دهد. او هریک از سیب‌زمینی‌ها را وزن کرده و جدول زیر را تهیه نمود :

تعداد سیب‌زمینی‌های موجود در کیسه	۲۰	۱۶	۱۸	۱۴	۱۴	۱۵	۱۳	۱۰
میانگین وزن سیب‌زمینی‌ها برحسب گرم	۷۲	۱۰۵	۱۱۰	۱۲۵	۱۳۶	۱۴۲	۱۷۴	۱۹۲

ــ نمودار پراکنش داده‌ها را رسم کنید (۱ سانتی‌متر روی محور xها نشان‌دهندهٔ یک سیب‌زمینی و ۱ سانتی‌متر روی محور yها نشان دهنده ۲۰ گرم)

ــ خط رگرسیون را رسم کنید.

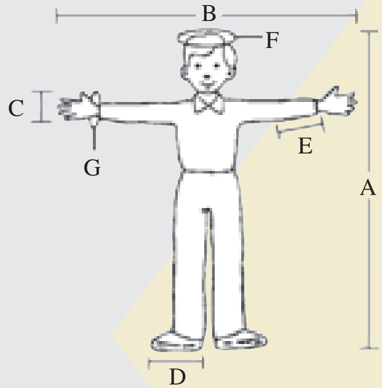
ــ ارتباط بین تعداد سیب‌زمینی‌های موجود در یک بسته را با میانگین وزن بیان کنید.

ــ با استفاده از نمودار، میانگین وزن سیب‌زمینی‌ها را در بسته‌ای که ۱۷ سیب‌زمینی وجود دارد، تخمین بزنید.

ــ در یک بسته ۲ کیلوگرمی اگر میانگین وزن ۱۶۰ گرم باشد، تعداد سیب‌زمینی‌های بسته را (به نزدیک‌ترین عدد صحیح) تخمین بزنید.

پروژه ۱

در اندازه‌گیری زیر، کدام‌یک از اندازه‌ها با هم مرتبط هستند.



A : قد

B : فاصله بین دو دست

C : وجب

D : طول پا

E : مچ تا آرنج

F : دور سر

G : دور مچ دست

ـ اندازه‌های فوق را برای چند تن از همکلاسی‌هایتان به‌دست آورید.

ـ با استفاده از اندازه‌های به‌دست آمده معین کنید آیا جفت متغیری وجود دارد که اندازه‌شان با هم مرتبط باشند.

اگر چنین است، خط رگرسیون را برایشان رسم کنید. گزارشی تهیه کنید و روی نتیجه نظر بدهید. آیا با آنچه قبل از شروع پروژه حدس زدید هماهنگی دارد یا خیر؟ اگر داده‌ها را از دو گروه زنان و مردان به‌دست آورید چه اتفاقی می‌افتد؟



پروژه ۲

می‌خواهیم بدانیم آیا ارتباطی بین موفقیت محصل در دروس ریاضی و دروس غیرریاضی وجود دارد؟ برای این منظور یکی از دروس ریاضی و یکی از دروس غیرریاضی خود را انتخاب کنید و با نمونه مناسبی از کلاس درس وجود نوع ارتباط را تحقیق کنید. ضریب تغییرات در کدام یک از این دروس بیش‌تر است . در صورت وجود ارتباط خطی معادله خط کم‌ترین مربعات را به‌دست آورید.

ــ نتایج کار خود را با نتایج به‌دست آمده از تحقیق گروه‌های دیگر مقایسه کنید. آیا نتایج به دست آمده یکسان است؟



پروژه

آیا بین اوقات فراغت دانش‌آموزان و
عملکرد آن‌ها در مدرسه ارتباطی وجود دارد؟

معیارهای ارزشیابی پروژه

هر پروژه با ۶ معیار ارزشیابی می‌شود :

معیار ۱ــ تعریف و طراحی پروژه: در این قسمت پروژه باید کاملاً تعریف شده باشد، طرحی برای اجرا و به پایان رسانیدن پروژه ارائه شده باشد.

۳ــ۰ امتیاز

به‌طور مثال در پروژه ضمیمه این کار در قسمت خلاصه و مقدمه انجام شده است. پروژه کاملاً تعریف شده و شمای کلی کار ارائه شده است.

معیار ۲ــ جمع‌آوری داده‌ها: در این قسمت روش جمع‌آوری داده‌ها، ارتباط داده‌های جمع‌آوری شده با موضوع پروژه، سازمان‌دهی داده‌های جمع‌آوری شده و کافی بودن داده‌ها از نظر کمّی و کیفی مورد نظر است.

۵ ــ۰ امتیاز

معیار ۳ــ تجزیه‌و تحلیل داده‌ها: در این قسمت تجزیه و تحلیل داده‌ها از دو جنبهٔ محاسبات و ارائه نموداری مورد ارزشیابی قرار می‌گیرد.

استفادهٔ مناسب از نمودارها، دقت در محاسبات عددی و استفاده از تکنیک‌های مختلف

تجزیه و تحلیل داده‌ها موردنظر می‌باشد.

۶ ــ ۰ امتیاز

معیار ۴ــ ارزیابی: در این قسمت نتیجه‌گیری و ارزیابی نتایج حاصل از جمع‌آوری و تجزیه‌وتحلیل داده‌ها و نیز نقد پروژه توسط خود دانش‌آموز مورد توجه است.

۶ ــ ۰ امتیاز

معیار ۵ــ برقراری ارتباط: چگونگی ارائهٔ پروژه و استفادهٔ صحیح از نمادها مورد بررسی می‌باشد.

۵ ــ ۰ امتیاز

معیار ۶ــ همکاری و ارائه شفاهی پروژه در کلاس

۱۵ ــ ۰ امتیاز

خلاصه

اغلب، دانش‌آموزان فکر می‌کنند، به‌دلیل آن که استراحت کافی ندارند، نمی‌توانند تلاش بیش‌تری در مطالعه داشته باشند. برخی فکر می‌کنند چرا همکلاسی‌هایشان نمرات بهتری کسب می‌کنند ولی آن‌ها نمی‌توانند نمرهٔ خوبی کسب کنند. برخی فکر می‌کنند اگر اوقات فراغت بیش‌تری داشتند می‌توانستند عمل‌کرد بهتری داشته باشند. سؤال این است که **آیا ارتباطی بین اوقات فراغت دانش‌آموزان و عمل‌کرد آن‌ها در مدرسه وجود دارد؟** این سؤال از طریق این پروژه مورد بررسی قرار گرفت.

نمونه آماری در این پروژه، دانش‌آموزانی هستند که به‌طور تصادفی از میان دانش‌آموزان مدارس بین‌الملل دختران و پسران تهران انتخاب شدند. با استفاده از محاسبات آماری، این پروژه به پایان رسید. نتیجه نشان داد که این ارتباط وجود ندارد.

فهرست محتوا

مقدمه

هدف این پروژه، بررسی ارتباط بین اوقات فراغت و عمل‌کرد دانش‌آموزان در مدرسه می‌باشد. برای یک دانش‌آموز کاملاً عادی است که بشنود : «چرا وقت خودت را تلف می‌کنی؟» «چرا بیش‌تر درس نمی‌خوانی؟» و یا نصیحت‌هایی چون «بیش‌تر درس بخوان».

بیش‌تر دانش‌آموزان دوست ندارند نصیحت‌های خانواده‌هایشان را گوش کنند ولی آیا این دانش‌آموزان به اندازهٔ کافی احساس مسئولیت می‌کنند و از وقت خود استفاده درست و لازم را دارند؟ آیا در برنامه‌ریزی برای اوقات خود معقول عمل می‌کنند؟ چقدر از زمان خود را صرف مطالعه می‌کنند و چقدر از آن را به‌هدر می‌دهند؟ ممکن است یک روز ۲ ساعت هدر شود، روز دیگر ۳ ساعت ... ولی در پایان هفته این ساعات باهم جمع می‌گردد. ساعاتی که می‌شد بهرهٔ بیش‌تری از آن‌ها جست و عمل‌کرد بهتری داشت.

بیش‌تر خانواده‌ها می‌گویند : شما به اندازه کافی درس نمی‌خوانی و اگر بیش‌تر درس بخوانی نمرهٔ بهتری خواهی گرفت. آیا چنین است؟

این پروژه با توزیع پرسش‌نامه بین دانش‌آموزان آغاز گشت. در این رابطه داده‌ها جمع‌آوری شده و مورد ارزیابی و تجزیه و تحلیل قرار گرفتند. در این بررسی از محاسبات آماری، نمودارها و نیز محاسبه احتمالات شرطی استفاده شد.

جمع‌آوری داده‌ها

به‌دلیل مشکلات موجود برای جمع‌آوری داده در مدارس، دو مدرسه بین‌الملل تهران به‌عنوان جامعهٔ آماری درنظر گرفته شد و ۴۰ دانش‌آموز به‌طور تصادفی از بین دانش‌آموزان این دو مدرسه به‌عنوان نمونه انتخاب شدند. دانش‌آموزان از **کلاس‌های اول دبیرستان تا پیش‌دانشگاهی** پسر و دختر، انتخاب شدند.

جمع‌آوری داده‌ها می‌توانست از دوطریق مصاحبه و پرسش‌نامه انجام شود ولی روش پرسش‌نامه به‌نظر مناسب‌تر آمد. در این نظرخواهی، اسامی خواسته نشده بود تا دانش‌آموزان بتوانند با راحتی خیال پاسخ‌های خود را بدهند. از طریق مصاحبه ممکن بود دانش‌آموزان پاسخ‌های درست ندهند و نیز زمان بیش‌تری مورد نیاز بود. در پرسش‌نامه از طول زمان و چگونگی گذراندن اوقات فراغت دانش‌آموزان سؤال شد و این که چه‌مقدار از این زمان فراغت صرف مطالعه می‌شود و معدل دانش‌آموزان چقدر است؟ یک نسخه از پرسش‌نامه و برخی جداول حاوی اطلاعات ضمیمه این گزارش است.

روند محاسبهٔ اوقات بیکاری نیز در ضمیمه C آمده است. جداول ۱ و ۲و ۳ و ۴ برخی از داده‌های جمع‌آوری شده را نشان می‌دهند.

جدول۱: زمان صرف‌شده برای فعالیت‌های غیر از درس‌خواندن در هفته

شمارهٔ دانش‌آموزان	زمان برحسب ساعت در هفته
۱	۹
۲	۷
۳	۱۳
۴	۶
۵	۱۸
۶	۲/۵
۷	۴
۸	۴/۵
۹	۴۹
۱۰	۳۲
۱۱	۱۷
۱۲	۱۷
۱۳	۳
۱۴	۵
۱۵	۶
۱۶	۷۶
۱۷	۹
۱۸	۲
۱۹	۷
۲۰	۲۲
۲۱	۳۰
۲۲	۳۵
۲۳	۳۰
۲۴	۲۴
۲۵	۴۵
۲۶	۳۰
۲۷	۷۱
۲۸	۲۰
۲۹	۸
۳۰	۱
۳۱	۶
۳۲	۷

جدول۲: زمان صرف‌شده برای مطالعه بعد از مدرسه در اوقات فراغت (در هفته)

شمارهٔ دانش‌آموزان	زمان برحسب ساعت در هفته
۱	۰
۲	۵
۳	۲۸
۴	۶
۵	۶
۶	۶ـ۵
۷	۳
۸	۳
۹	۴
۱۰	۱۴
۱۱	۱۶ـ۱۴
۱۲	۴۵
۱۳	۵
۱۴	۸
۱۵	۱۲
۱۶	۱۰
۱۷	۳
۱۸	۸ـ۷
۱۹	۲۸
۲۰	۱۵ـ۱۲
۲۱	۲
۲۲	۲
۲۳	۲
۲۴	۲
۲۵	۲
۲۶	۱۴
۲۷	۴
۲۸	۴
۲۹	۸
۳۰	۱۰
۳۱	۱۰
۳۲	۱۲

جدول۳: معدل آخرین سال دانش‌آموزان

شمارهٔ دانش‌آموزان	معدل برحسب درصد
۱	۹۵
۲	۹۶/۷۵
۳	۹۷
۴	۸۸
۵	۸۶
۶	۸۳
۷	۹۴/۵
۸	۸۸/۶۳
۹	۷۰
۱۰	۸۶
۱۱	۸۹/۸
۱۲	۷۸/۵
۱۳	۷۱
۱۴	۸۶
۱۵	۷۸/۵
۱۶	۸۵
۱۷	۸۰
۱۸	۸۳
۱۹	۸۵
۲۰	۹۳
۲۱	۹۳
۲۲	۹۳
۲۳	۵۹/۵
۲۴	۶۵/۴
۲۵	۷۳/۵
۲۶	۷۸/۵
۲۷	۷۳/۵
۲۸	۷۳/۶
۲۹	۷۸
۳۰	۸۰
۳۱	۷۸
۳۲	۷۰

جدول۴: زمان فراغت دانش‌آموزان

شمارهٔ دانش‌آموزان	ساعات فراغت
۱	۵۰/۷۵
۲	۳۵
۳	۴۲
۴	۳۹
۵	۵۰
۶	۳۳/۸۳
۷	۲۱
۸	۱۵
۹	۲۸
۱۰	۳۹/۵
۱۱	۳۹/۵
۱۲	۵۳/۵
۱۳	۶۳/۱۶
۱۴	۵۶/۵
۱۵	۵۹/۵
۱۶	۵۶/۸۳
۱۷	۳۴
۱۸	۴۶
۱۹	۳۷/۵
۲۰	۴۳
۲۱	۳۵/۵۸
۲۲	۳۸/۲۵
۲۳	۲۸/۵۸
۲۴	۵۲
۲۵	۲۵/۵
۲۶	۲۲/۵
۲۷	۳۸
۲۸	۲۸/۱۶
۲۹	۳۲/۵
۳۰	۳۱/۷۵
۳۱	۳۲/۵
۳۲	۲۵/۵

تجزیه و تحلیل داده‌ها

این قسمت از دو بخش محاسبات و نمودارها تشکیل شده است : ابتدا میانگین و انحراف معیار برای زمان فراغت و معدل هریک از دانش‌آموزان حساب شد. نمودارهای میله‌ای نیز رسم شدند.

صفحات ۱۹۲، ۱۹۳ و ۱۹۴.

ارزیابی مشکلات

در انجام این پروژه متأسفانه با مشکلاتی روبرو بودم که عبارتند از :

ــ به دلیل مشکل عدم آشنایی به زبان فارسی این بررسی محدود به مدارس بین‌الملل تهران شد.

ــ در این بررسی ۸ پاسخ‌نامه به دلیل مخدوش بودن پاسخ‌ها کنار گذاشته شد. بنابراین نمونه به ۳۲ کاهش یافت.

ــ محدودیت زمان

نتایج

قبل از آن که هرگونه نتیجه گیری شود، نتایج مشاهدات به قرار زیر ارائه می‌شوند.

ــ محاسبات آماری روی معدل نشان دادند که ۷ نفر از ۳۲ نفر دانش‌آموزان عمل‌کرد بسیار عالی داشتند. ۱۱ نفر عمل‌کرد خوب و ۹ نفر عمل‌کرد قابل قبول ، ۴ نفر ضعیف و یک نفر خیلی ضعیف بودند.

ــ محاسبات روی اوقات فراغت نشان دادند که ۶ نفر وقت آزاد بسیار داشتند و ۷ نفر بیش‌تر از کافی، ۱۴ نفر وقت آزاد کافی داشتند. ۴ نفر وقت کافی نداشتند و یک نفر اصلاً وقت آزاد نداشت.

ــ اطلاعات اضافی عبارتند از ۲۴ نفر از ۳۲ نفر از اوقات بیکاری خود خسته می‌شوند.

ــ ۹ نفر از ۳۱ نفر حداقل یک‌بار در طول تحصیل‌شان از یک درس نمره قبولی نیاورده‌اند.

ــ ۲۶ نفر از ۳۰ نفر از عمل‌کرد خود راضی نیستند و می‌خواهند آن را بهبود دهند.

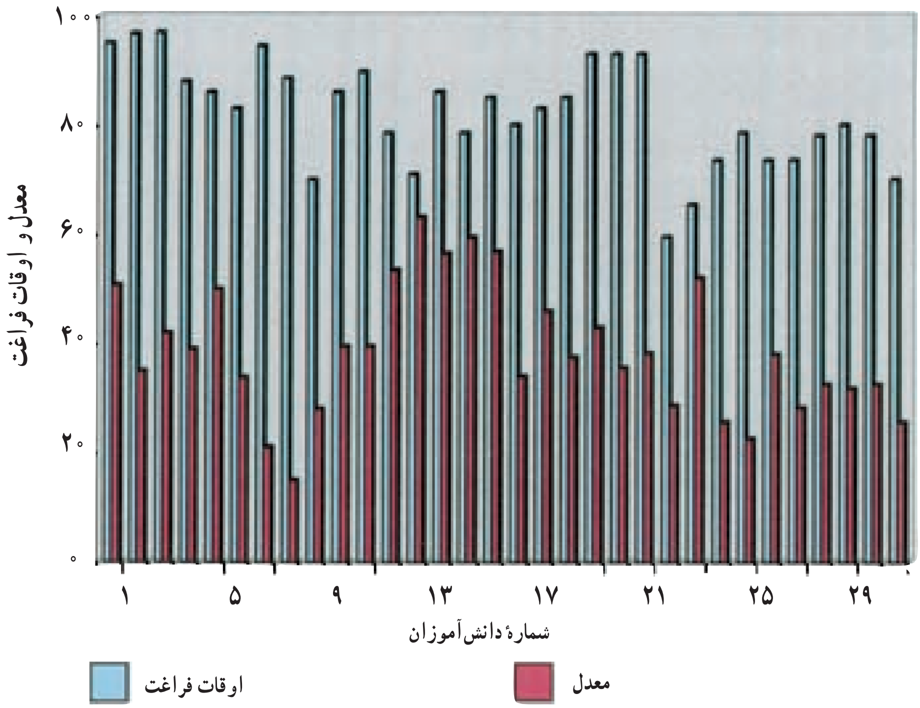
ــ۲۴ نفر از ۳۲ نفر عقیده داشتند که اگر وقت بیش‌تری صرف مطالعه کنند نتیجه بهتری خواهند داشت.

جدول ۱ــ جدول فراوانی اوقات بیکاری

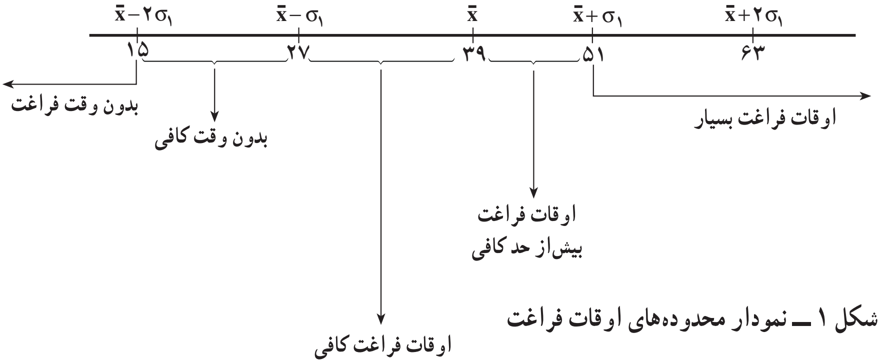
زمان فراغت	فراوانی	فراوانی× زمان فراغت	انحراف از میانگین	مجذور انحراف از میانگین	فراوانی× مجذور انحراف از میانگین
۱۵	۱	۱۵	-۲۴	۵۷۶	۵۷۶
۲۱	۱	۲۱	-۱۵	۲۲۵	۲۲۵
۲۲/۵	۱	۲۲/۵	-۱۶/۵	۲۷۲/۲۵	۲۷۲/۲۵
۲۵/۵	۲	۵۱	-۱۳/۵	۱۸۲/۲۵	۳۶۴/۵
۲۸	۱	۲۸	-۱۱	۱۲۱	۱۲۱
۲۸/۱۶	۱	۲۸/۱۶	-۱۰/۸۴	۱۱۷/۵۰۵۶	۱۱۷/۵۰۵۶
۲۸/۵۸	۱	۲۸/۵۸	-۱۰/۴۲	۱۰۸/۵۷۶۴	۱۰۸/۵۷۶۴
۳۱/۷۵	۱	۳۱/۷۵	-۷/۲۵	۵۲/۵۶۲۵	۵۲/۵۶۲۵
۳۲/۵	۲	۶۵	-۶/۵	۴۲/۲۵	۸۴/۵
۳۳/۸۳	۱	۳۳/۸۳	-۵/۱۷	۲۶/۷۲۸۹	۲۶/۷۲۸۹
۳۴	۱	۳۴	-۵	۲۵	۲۵
۳۵	۱	۳۵	-۴	۱۶	۱۶
۳۵/۵۸	۱	۳۵/۵۸	-۳/۴۲	۱۱/۶۹۶۴	۱۱/۶۹۶۴
۳۷/۵	۱	۳۷/۵	-۱/۵	۲/۲۵	۲/۲۵
۳۸	۱	۳۸	-۱	۱	۱
۳۸/۲۵	۱	۳۸/۲۵	-۰/۷۵	۰/۵۶۲۵	۰/۵۶۲۵
۳۹	۱	۳۹	۰	۰	۰
۳۹/۵	۲	۷۹	۰/۵	۰/۲۵	۰/۵
۴۲	۱	۴۲	۳	۹	۹
۴۳	۱	۴۳	۴	۱۶	۱۶
۴۶	۱	۴۶	۷	۴۹	۴۹
۵۰	۱	۵۰	۱۱	۱۲۱	۱۲۱
۵۰/۷۵	۱	۵۷/۷۵	۱۱/۷۵	۱۳۸/۰۶۲۵	۱۳۸/۰۶۲۵
۵۲	۱	۵۲	۱۳	۱۶۹	۱۶۹
۵۳/۵	۱	۵۳/۵	۱۴/۵	۲۱۰/۲۵	۲۱۰/۲۵
۵۶/۵	۱	۵۶/۵	۱۷/۵	۳۰۶/۲۵	۳۰۶/۲۵
۵۶/۸۳	۱	۵۶/۸۳	۱۷/۸۳	۳۱۷/۹۰۸۹	۳۱۷/۹۰۸۹
۵۹/۵	۱	۵۹/۵	۲۰/۵	۴۲۰/۲۵	۴۲۰/۲۵
۶۳/۱۶	۱	۶۳/۱۶	۲۴/۱۶	۵۸۳/۷۰۵۶	۵۸۳/۷۰۵۶
	۳۲	۱۲۳۴/۳۹			۴۳۴۶/۰۵۹۳

جدول ۲ــ جدول فراوانی برای معدل

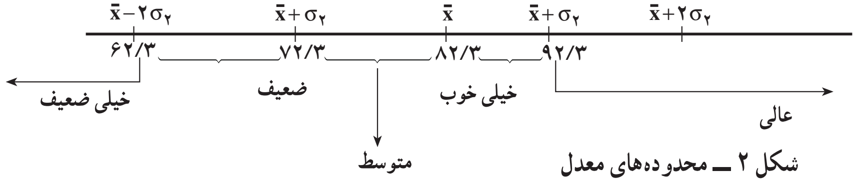
معدل	فراوانی	معدل×فراوانی	انحراف از میانگین	مجذور انحراف از میانگین	مجذور انحراف از میانگین× فراوانی
۶۰	۱	۶۰	-۲۳	۵۲۹	۵۲۹
۶۵	۱	۶۵	-۱۸	۳۲۴	۳۲۴
۷۰	۲	۱۴۰	-۱۳	۱۶۹	۳۳۸
۷۱	۱	۷۱	-۱۲	۱۴۴	۱۴۴
۷۴	۳	۲۲۲	-۹	۸۱	۲۴۳
۷۶	۱	۷۶	-۷	۴۹	۴۹
۷۸	۲	۱۵۶	-۵	۲۵	۵۰
۷۹	۳	۲۳۷	-۴	۱۶	۴۸
۸۰	۲	۱۶۰	-۳	۹	۱۸
۸۳	۲	۱۶۶	۰	۰	۰
۸۵	۲	۱۷۰	۲	۴	۸
۸۶	۳	۲۵۸	۳	۹	۲۷
۸۸	۱	۸۸	۵	۲۵	۲۵
۸۹	۱	۸۹	۶	۳۶	۳۶
۹۰	۱	۹۰	۷	۴۹	۴۹
۹۳	۳	۲۷۹	۱۰	۱۰۰	۳۰۰
۹۵	۳	۲۸۵	۱۲	۱۴۴	۴۳۲
۹۷	۲	۱۹۴	۱۴	۱۹۶	۳۹۲
	۳۴	۲۸۱۶			۳۰۱۲



ارتباط بین معدل و اوقات فراغت



شکل ۱ ــ نمودار محدوده‌های اوقات فراغت



شکل ۲ ــ محدوده‌های معدل

نتیجه‌گیری

با توجه به نمودارها و محاسبات انجام شده نتیجه می‌شود که ارتباط محسوسی بین **مقدار** اوقات فراغت و عمل‌کرد در مدرسه وجود ندارد. معدل دانش‌آموز ارتباطی به میزان اوقات فراغت ندارد. دانش‌آموزان بسیاری وجود دارند که وقت فراغت بسیار دارند ولی عمل‌کرد بالایی ندارند.

ضمیمه A: «پرسش‌نامه»

این پرسش‌نامه شامل سؤالاتی در زمینهٔ چگونگی گذراندن اوقات فراغت شما، زمانی که صرف فعالیت‌های غیر مدرسه‌ای و زمانی که صرف مطالعه می‌شود می‌باشد.

هدف این بررسی، پیدا کردن ارتباط بین اوقات فراغت دانش‌آموزان و عمل‌کرد آن‌ها در مدرسه است. این پرسش‌نامه ناشناس است. بنابراین، لطفاً صادقانه پاسخ دهید.

با تشکر از همکاری شما

۱ــ ساعت شروع مدرسه و پایان مدرسه را بیان کنید (لازم به ذکر است در این دو مدرسه ساعات درسی به دلیل انتخاب واحدها متغیر است)

	شنبه	یکشنبه	دوشنبه	سه‌شنبه	چهارشنبه	پنج‌شنبه
ساعت شروع مدرسه						
ساعت پایان مدرسه						

۲ــ چقدر زمان لازم است تا از مدرسه به‌خانه برگردید؟

۳ــ معمولاً ساعت چند می‌خوابید؟

۴ــ چند ساعت بعد از مدرسه در هفته کار می‌کنید؟

۵ــ چند ساعت در هفته به تماشای تلویزیون یا انجام دیگر تفریحات می‌پردازید؟

۶ــ معدل شما در سال گذشته چند است؟

۷ــ آیا از درسی مردود شده‌اید؟

در این ضمیمه اطلاعات مربوط به ۲۰ دانش‌آموز آورده شده است.

ضمیمهٔ B:

جدول ۱: ساعات شروع و پایان مدرسه در طول روزهای هفته

شمارهٔ دانش‌آموزان	شنبه		یکشنبه		دوشنبه		سه‌شنبه		چهارشنبه		پنجشنبه	
	شروع	پایان	شروع	پایان	شروع	پایان	شروع	پایان	شروع	پایان	شروع	پایان
۱	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۳,۱۵	ق.ظ ۸,۰۰	ب.ظ ۳,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	....	....
۲	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰
۳	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	....	....
۴	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	....	....
۵	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۴,۰۰	ق.ظ ۸,۰۰	ب.ظ ۴,۰۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۴,۰۰	ق.ظ ۸,۰۰	ب.ظ ۱۲,۳۰
۶	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۴,۰۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۴,۰۰	ق.ظ ۸,۰۰	ق.ظ ۱۱,۱۰
۷	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰
۸	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰
۹	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	....	....
۱۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	....	....
۱۱	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	....	....
۱۲	ق.ظ ۸,۰۰	ب.ظ ۱۲,۳۰	ق.ظ ۱۱,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۴,۰۰	ق.ظ ۸,۰۰	ب.ظ ۱۲,۳۰	ق.ظ ۹,۳۰	ب.ظ ۴,۰۰	ق.ظ ۸,۰۰	ب.ظ ۱۲,۳۰
۱۳	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۴,۰۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۱۱,۱۰	ب.ظ ۱۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۱۲,۳۰
۱۴	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۴,۰۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۴,۰۰	ق.ظ ۸,۰۰	ب.ظ ۱۲,۳۰
۱۵	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۴,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۹,۳۰	ب.ظ ۴,۳۰	ق.ظ ۸,۰۰	ب.ظ ۱۲,۳۰
۱۶	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	....	....
۱۷	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۳,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	....	....
۱۸	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۳,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	....	....
۱۹	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰
۲۰	ق.ظ ۸,۰۰	ب.ظ ۴,۰۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۹,۱۵	ب.ظ ۱۲,۳۰	ق.ظ ۸,۰۰	ب.ظ ۲,۳۰	ق.ظ ۹,۱۵	ب.ظ ۴,۰۰	ق.ظ ۸,۰۰	ق.ظ ۱۱,۰۰

*ــ ق. ظ یعنی قبل از ظهر و ب. ظ یعنی بعدازظهر

جدول۳: زمان لازم برای رسیدن از خانه به مدرسه

شمارهٔ دانش‌آموزان	زمان (ساعت)
۱	۰/۳۰
۲	۰/۳۰
۳	۰/۳۰
۴	۳/۳۰
۵	۰/۳۰
۶	۰/۴۵
۷	۰/۳۰
۸	۱/۳۰
۹	۰/۳۰
۱۰	۱/۰
۱۱	۱/۰
۱۲	۰/۱۵
۱۳	۰/۳۰
۱۴	۰/۳۰
۱۵	۰/۳۰ـ۱/۰
۱۶	۰/۲۰
۱۷	۰/۲۰ـ۰/۳۰
۱۸	۰/۱۵ـ۰/۳۰
۱۹	۰/۰۵
۲۰	۰/۳۵
۲۱	۰/۳۰
۲۲	۰/۴۰
۲۳	۰/۳۰
۲۴	۰/۳۰
۲۵	۰/۳۰
۲۶	۱/۳۰
۲۷	۰/۴۵
۲۸	۱/۳۰
۲۹	۱/۰
۳۰	۰/۴۵
۳۱	۱/۰
۳۲	۱/۰

جدول۴: ساعت خواب دانش‌آموزان

شمارهٔ دانش‌آموزان	زمان (ساعت)
۱	قبل از ظهر ۱۲,۳۰
۲	بعد از ظهر ۱۱,۰۰
۳	بعد از ظهر ۱۱,۰۰
۴	قبل از ظهر ۱۲,۰۰
۵	قبل از ظهر ۱,۳۰
۶	بعد از ظهر ۱۱,۰۰
۷	بعد از ظهر ۹,۰۰
۸	بعد از ظهر ۹,۰۰
۹	بعد از ظهر ۱۰,۰۰
۱۰	بعد از ظهر ۱۱,۰۰
۱۱	بعد از ظهر ۱۱,۰۰
۱۲	قبل از ظهر ۱,۳۰
۱۳	قبل از ظهر ۲,۰۰
۱۴	قبل از ظهر ۲,۰۰
۱۵	قبل از ظهر ۳,۰۰
۱۶	قبل از ظهر ۱,۰۰
۱۷	بعد از ظهر ۱۰,۰۰
۱۸	قبل از ظهر ۱۲,۰۰
۱۹	بعد از ظهر ۱۱,۰۰
۲۰	بعد از ظهر ۱۱,۳۰
۲۱	بعد از ظهر ۱۱,۰۰
۲۲	بعد از ظهر ۱۱,۰۰
۲۳	بعد از ظهر ۱۰,۰۰
۲۴	بعد از ظهر ۱۱,۰۰
۲۵	بعد از ظهر ۹,۳۰
۲۶	بعد از ظهر ۱۰,۰۰
۲۷	بعد از ظهر ۱۱,۰۰
۲۸	بعد از ظهر ۱۰,۳۰
۲۹	بعد از ظهر ۱۰,۰۰
۳۰	بعد از ظهر ۱۰,۰۰
۳۱	بعد از ظهر ۱۰,۰۰
۳۲	بعد از ظهر ۹,۰۰

ضمیمه C: محاسبهٔ اوقات فراغت

اوقات فراغت دانش‌آموزان به صورت زیر محاسبه شده است. طول روز ۱۲ ساعت درنظر گرفته شده است. ساعت در هفته ۸۴ = ۱۲×۷

محاسبهٔ اوقات فراغت دانش‌آموز شماره ۱ بنابر جدول (ضمیمه B): تعداد ساعاتی را که دانش‌آموزان در مدرسه صرف می‌کنند می‌توان پیدا کرد : به‌طور مثال :

ساعت در هفته ۳۴/۲۵ = ۶/۵+۶/۵+۷/۵+۷/۲۵+۶/۵

با استفاده از جدول ۲ (ضمیمه B) ، زمان لازم برای رسیدن به منزل در طول هفته عبارت است از :

روز ساعت

ساعت در هفته ۲/۵ = ۵×۰/۵

دانش‌آموز شماره ۱ معمولاً ساعت نیم بعد از نیمه شب می‌خوابد پس طول روز برای او ۵ ساعت در روز بیش‌تر از فرض ۱۲ ساعت است پس ۳/۵ = ۷×۰/۵ باید به زمان فراغت اضافه شود :

ساعت ۵۹/۷۵ = ۳/۵+[(۸۴−(۳۴/۵+۲/۵))]

وقت بیکاری دانش‌آموز شماره ۱

منابع و مراجع

۱ــ سالنامه آمار کشور، سال‌های ۱۳۷۱ و ۱۳۷۳ و ۱۳۷۶، مرکز آمار ایران

2- DUNCAN AND CHRISTINE GRAHAM, MAINSTREAM MATHEMATICS FOR GCSE, MACMILLAN PRESS ltd.

3- GEOFF BUCKWELL. GCSE CORE MATHEMATICS, MCMILLAN WORKOUT SERIES.

4- ALLAN G. BLUMAN, ELEMEN TARY STATISTICS, Wm .C. BROWN PUBLISHER.

5- PERM S. MANN, INTRODUCTORY STATISTICS, JOHN WILEY & SONS.

6-BRASE & BRASE, UNDERSTANDABLE STATIS TIES.

7- GEOFF BUCKWELL, MASTERING MATHEMATICS, MCMILLAN MASTER SERIES.

8- HAROLD R. JACOBS, MATHEMATICS A HUMAN ENDEAVOR, FREEMAN.

9- JAN DE LANGE JZN, HELEEN VERHANGE, DATA VISUALIZATION.

10- EWART SMITH, MATHS FOR INTERMEDIATE GNVQ.

11- B.L. AGARWAL, BASIC STATISTICS.

12- JAMES M. LANDWEHR & ANN E. WATKINS EXPLORING DATA.

13- ROBERT JOHNSON, ELEMENTARY STATISTICS, SIXTH EDITION DUXBURY PRESS.

14- JOHN E. FREUND, GARY A. SIMON, STATISTICS A FIRST COURSE FIFTH EDITION. PRENTICE - HALL INTERNATIONAL EDITIOW.

15- THE SCHOOL MATEMATICS PROJECT. STATISTICS, MATHEMATICS FOR A AND AS LEVEL, CAMBRIDGE UNIVERSITY PRESS.

16- MATHEMATICS PLUS. HARCOURT BRACE & COMPANY.

17- HOUGHTON MIFFLIN, UNIFIED MATHEMATICS, HOUGHTON MIFFLIN COMPANY / BOSTON.

18- SUE BRIGGS, HIGHER GCSE, OXFORD MATHEMATICS, OXFORD UNIVERSITY PRESS.

تعدادی از تصاویر از مجموعه مجلات NATIONAL GEOGRAPHIC انتخاب شده‌اند.

معلمان محترم، صاحب نظران، دانش آموزان عزیز و اولیای آنان می توانند نظر اصلاحی خود را درباره‌ی مطالب این کتاب از طریق نامه به نشانی تهران ـ صندوق پستی ۱۵۸۵۵/۳۶۳ ـ گروه درسی مربوط و یا پیام‌نگار (Email) talif@talif.sch.ir ارسال نمایند.

دفتر برنامه‌ریزی و تألیف کتاب‌های درسی